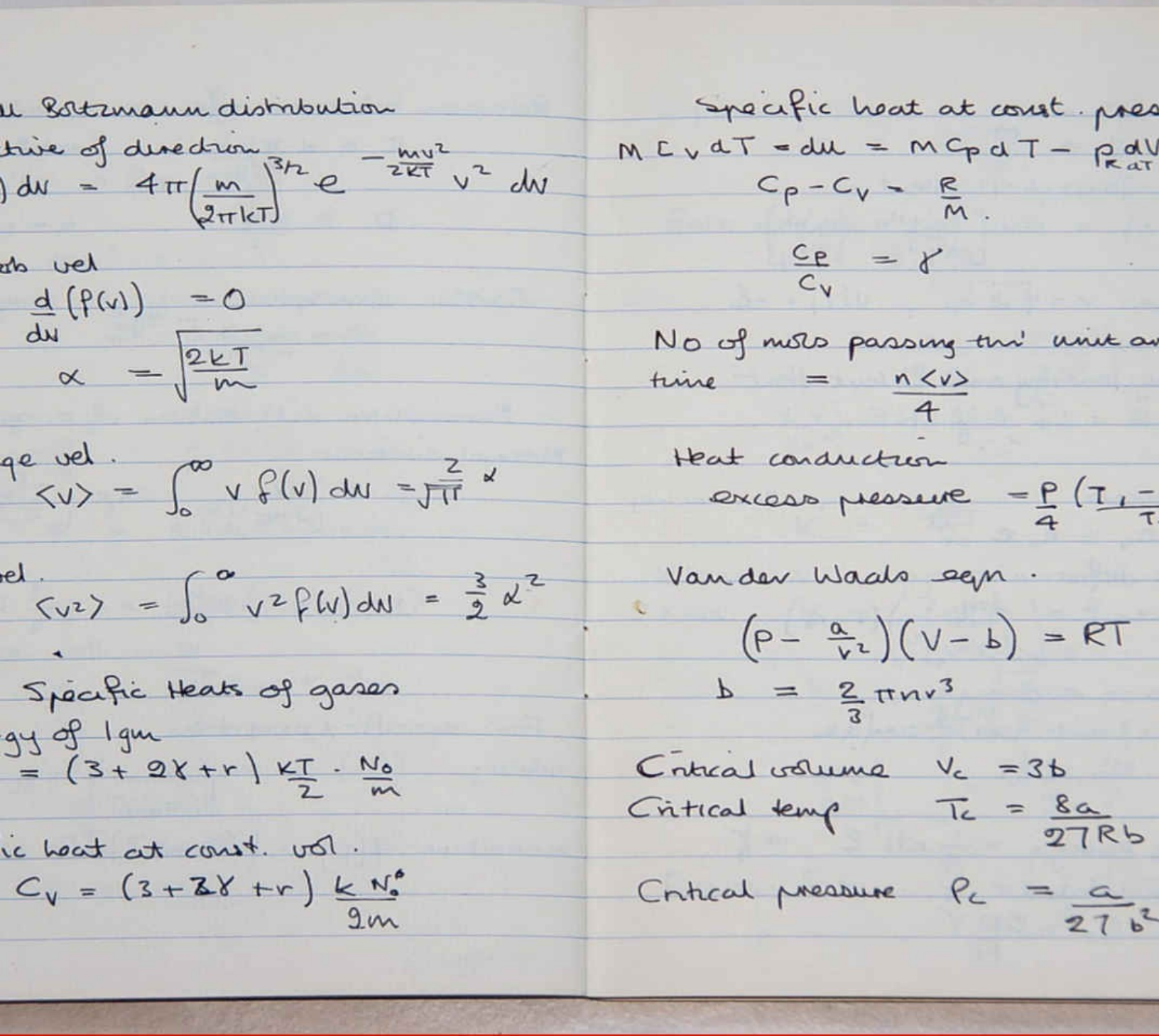

ll Boltzmann distribution
tive of direction
$) dN = 4\pi \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-\frac{mv^2}{2kT}} v^2 dN$

ob vel
$\frac{d}{dN}(f(v)) = 0$
$\alpha = \sqrt{\frac{2kT}{m}}$

ge vel.
$\langle v \rangle = \int_0^\infty v f(v) dN = \frac{2}{\sqrt{\pi}} \alpha$

el.
$\langle v^2 \rangle = \int_0^\infty v^2 f(v) dN = \frac{3}{2} \alpha^2$

Specific Heats of gases
gy of 1gm
$= (3 + 2\gamma + r) \frac{KT}{2} \cdot \frac{N_0}{m}$

ic heat at const. vol.
$C_V = (3 + 2\gamma + r) \frac{k N_0}{2m}$

Specific heat at const. pres
$M C_V dT = dU = M C_p dT - p_R dV$
$C_p - C_V = \frac{R}{M}$.

$\frac{C_p}{C_V} = \gamma$

No of mols passing thr' unit ar
time $= \frac{n \langle v \rangle}{4}$

Heat conduction
excess pressure $= \frac{p}{4} (T_1 -$

Van der Waals eqn.
$\left(p - \frac{a}{v^2}\right)(V - b) = RT$
$b = \frac{2}{3} \pi n r^3$

Critical volume $V_c = 3b$
Critical temp $T_c = \frac{8a}{27Rb}$
Critical pressure $P_c = \frac{a}{27 b^2}$

LOGICO-LINGUISTIC
PAPERS
BY J.-M. KUCZYNSKI

Logico-Linguistic Papers

J.-M. Kuczynski

Introduction

Slinging the slingshot

Implicit Comparatives and the Sorites

The Ravens

Are Definite Descriptions Referring Terms?

Non-declarative contexts and the theory of descriptions

Counterfactuals: The Epistemic Analysis

Sentential Reference

Sentence-reference versus Sentence-meaning

Sentential Reference the Foundations of Semantics

Slippery Slope Paradoxes

The Sorites and Pseudo-monadic Predicates

The concept of a symbol and the vacuousness of

the symbolic conception of thought

Grammatical inflexions in connection with the nature of Sentence-level Meaning

A Post-Kripkean Analysis of Analyticity

Extensionality

Knowledge of Referent versus of Knowledge of Meaning

Quantification in Natural Language

Intensionality

Analyticity

A cure for Nathan Salmonella: How to do semantics without butchering the laws of logic

No Such thing as a Meaningless Sentence: Erin McCready Refuted

A Review of Erin McCready's Review of

Conceptual Atomism and the Computational Theory of Mind

Induction by enumeration vs. inference to the best explanation

Introduction

In this volume, clear answers are given to age-old questions relating to:

*The nature of counterfactuals
*Extensionality
*The nature of linguistic meaning and the distinction between meaning and reference
*The relationship between logical form and grammatical form
*Analytic truth
*The relationship between cognitive content (what we think) and linguistic meaning (what we say)
*Hempel's Paradox (a.k.a. the Raven Paradox)
*The Paradox of the Heap (a.k.a. the Sorites Paradox), and finally
*The nature, and justification of, of inductive inference.

Slinging the slingshot

1.0 The Slingshot Consider the following sentences:

(1) Mozart was a composer.
(2) The class of all things x such that Mozart was a composer and x is self-identical is identical with the class of all things x such that x is self-identical.

(3) The class of all things x such that Bill Clinton was U.S. President in 1997 and x is self-identical is identical with the class of all things x such that x is self-identical.
(4) Bill Clinton was U.S. President in 1997.

Here is the position of those who believe in the cogency of so-called "slingshot"-style arguments:

(SS) **(1) is logically equivalent with (2). Therefore, (1) and (2) must co-refer, supposing that either refers to anything. (It is being assumed that logically equivalent expressions cannot have different referents.)**

(3) is what results when a referring term in (2) is replaced with a co-referring term.) ("The class of all things x such that Mozart was a composer and x is self-identical" refers to the class of all self-identical objects, and so too does "the class of all things x such that Bill Clinton was U.S. President in 1997 and x is self-identical.") Since intersubstituting co-referring terms preserves reference, it follows that (2) and (3) co-refer, supposing that either refers to anything.

(3) is equivalent with (4). Therefore, (3) and (4) must co-refer, supposing that either refers to anything.

Therefore, (1) and (4) must co-refer, supposing that they refer to anything.

If the occurrences in (1)-(4) "Mozart was a composer" and "Bill Clinton was U.S. President in 1997" are replaced with occurrences of *any* two true sentences, the argument that results is cogent. Thus, any two true sentences co-refer, supposing that any true sentence refers to anything. (For the same reason *mutatis mutandis*, any two false sentences co-refer, supposing that any false sentence refers to anything.)

The property of being true is the one semantically significant thing that (1) and (4) have in common. So, supposing that sentences refer to anything, (1) and (4) must refer to the property of being true, and true sentences must in general be assumed to refer to the property of being true.[1]" For the same reason *mutatis mutandis*, false sentences must refer to the property of being false.[2]

In general, a slingshot-style argument is one to the effect that, since logically equivalent expressions can't have different referents and since intersubstituting co-referring terms preserves truth-value, some class that would seem to comprise several objects (for example, the class of things that are referred by true sentences) in fact only comprises one object (for example, the property of being true).

1.1 One problems with SS and with other slingshot-style arguments SS is a poor argument that embodies a spurious understanding of reference, the same being true of every other slingshot-style argument. Suppose for argument's sake that "the class of all things x such that Mozart was a composer x is self-identical" is functioning as a device of reference. In that case, it's picking out the class of self-identical objects, and what (2) is saying about that class is that it's identical with itself; in other words, (2), thus interpreted, says that:

(2*) the class of self-identical objects is identical with the class of self-identical objects.

But (2*) is not equivalent with (1). Therefore, if the definite descriptions in

(1) and (2) are assumed to be referring terms—which is precisely what they *are* being assumed to be—(2) isn't equivalent with (1). For the same reason *mutatis mutandis*, (4) is not equivalent with (3) if "the class of all things x such that Bill Clinton was U.S. President in 1997 and x is self-identical" is taken to be a device of reference.

In order for (2) to be equivalent with (1), the former must be taken to be to the effect that:

(2#) there exists a class to which a thing belongs iff Mozart is a composer and that thing is self-identical; and there exists a class to which a thing belongs iff that thing is self-identical; and the first class is identical with the second class.

But if (2) is interpreted in this way, the definite description occurring in it is a quantifier[3], not a referring term. And, in that case, (3) is *not* what result when a referring term is replaced with a co-referring term. (It's what results when a non-referring term is replaced with a non-referring term.)

So SS is a failure.

Note: Some would disagree with our contention that, when read attributively, occurrences of definite descriptions are functioning as quantifiers, not as referring terms. Some, in other words, hold that there is such a thing as "attributive" reference. This is the position of Frege and his followers (e.g. Church). In Sections 2.0-2.3, we'll see why Frege's position is both indefensible and unnecessary.

1.2 Is the conclusion of SS correct? Do true sentences refer to truth-values? In this section we'll see that, in respect of its semantic properties, "snow is white" bears little resemblance to "the property of being true." And

we'll see that, for this reason, it is implausible, if not incoherent, to hold that "snow is white" (or any other true sentence) co-refers with "the property of being true" (or "*the True*" or "the truth-value *True*").[4]

Let's start with some platitudes about reference. To say that expression R refers to object O is to say that, for any property phi, ⌜R has phi⌝ attributes phi to O.[5] If you want to say of Benjamin Franklin (BF) that he's bald, you can do so by saying "Benjamin Franklin is bald." (Of course, BF is no longer with us. But to set the aside some extremely intricate and, in this context, irrelevant issues relating to tense, let us pretend that Franklin is alive and that we can speak of him in the present tense.) And (the remaining semantic rules of English being what they are) if you *couldn't* attribute baldness to BF by uttering that sentence, then "Benjamin Franklin" wouldn't refer to BF.[6] If, in saying "Benjamin Franklin is bald," one was attributing baldness to *Aristotle*, then, in that context at least, "Benjamin Franklin" would refer to Aristotle.

Let's generalize these points. For any property phi, "Benjamin Franklin" refers to BF (and not to Aristotle, for example) because ⌜Benjamin Franklin has phi⌝ attributes phi to Benjamin Franklin; and if one *couldn't* attribute phi to Benjamin Franklin by saying ⌜Benjamin Franklin has phi⌝, then "Benjamin Franklin" would not refer to BF. Further, if, in uttering that sentence, one was attributing baldness to Aristotle (but not BF), then "Benjamin Franklin" *would* refer to Aristotle (but not Benjamin Franklin). Thus, for "Benjamin Franklin" to refer to BF it is both *necessary* and *sufficient* that ⌜Benjamin Franklin has phi⌝ attribute phi to BF. The fact that "Benjamin Franklin" refers to BF is *identical* with the fact that ⌜Benjamin Franklin has phi⌝ attributes phi to Benjamin Franklin, for any property phi.

In general, expression E refers to O if and *only* if, for any property phi, ⌜E has phi⌝ attributes phi to O. E's referring to O may be *identified* with E's being such that, in virtue of containing it, a sentence *ipso facto* attributes a property to O.

Bearing these points in mind, let's revisit the question: Do sentences refer to truth-values? Does "snow is white" refer to the property of being true (which, in this context, we're assuming for argument's sake to be identical with the truth-value *true*)? No. In virtue of having the form, ⌜…the property of being true…⌝, a sentence is about the property of being true. ("Everything Bill says has the property of being true" says of the property of being true that it is instantiated by everything that Bill says.") But it isn't the case that, in virtue of having the form ⌜…snow is white…⌝, a sentence is about the property of being true.

Of course, "if snow is white, then snow isn't green" can, like any other sentence, be interpreted as concerning the property of being true—it can be interpreted as saying that the property of being true is a characteristic of the proposition that snow isn't green if it's white. But given how artificial it is to paraphrase "if snow is white, then snow isn't green" in this way, it's clear that, in some respect that is deeply relevant to their semantics, "the property of being true" and "snow is white" aren't comparable. Which, in its turn, suggests that it isn't *in virtue of* it's containing an occurrence of "snow is white" that "if snow is white, then snow isn't green" concerns that property.

But even if, in virtue of having the form ⌜…snow is white…⌝, a sentence *is* about the property of being true, the *way* in which, in virtue of having that form, such a sentence concerns that property is obviously different from the *way* in which, in virtue of having the form ⌜…the property of being true…⌝, a sentence concerns the property of being true. A story may help make this clear. You are my philosophy professor and I am taking an exam that consists

of just one question: I must identify at least one property that all propositions of the form *x is even iff x is divisible by two* have in common. I duly say (or write) "the property of being true." I pass. But I would have failed if I had instead said "snow is white" (or "grass is green"). In any case, you'd have no reason *independently of Frege's argument* for passing me. Which, taken together with the points made in the previous section, suggests that you'd have no reason at all for passing me.

It's hard to find any reason, apart from Frege's argument to the contrary, to hold that "snow is white" refers to the property of being true. Which, given how far Frege's argument to the contrary falls short of being probative, suggests that "snow is white" does not so refer.

1.2.1 Can sentences coherently be stipulated to have truth-values for their referents? Some (e.g. Church 1943) have *stipulated* that true sentences refer to that property (and that false ones refer to the property of being false). But some stipulations are incoherent, and this particular stipulation falls into that category. If I stipulate that "Gloxo" is to unambiguously refer to the number two, I cannot coherently stipulate that it is also to refer to the number of sides had by an arbitrary pentagon.

Bearing this in mind, consider the following variant of the just-told story. You are my geography professor, and you've just given me a test containing just one question: identify the state that lies on the Northern border of California. Instead of saying "Oregon", which is the right answer, I say "snow is white." You duly fail me.

Of course, if you and I have agreed that, by our stipulation, "snow is white" is to refer to Oregon, then "snow is white" would be an acceptable answer. But *to the extent that* "snow is white" would, for that reason, have been an acceptable answer, it obviously wouldn't have been functioning in

the same way as the homonymous *sentence* "snow is white." And it is thus incoherent to stipulate that "snow is white" is *unambiguously* both an English sentence (to the effect that a certain crystalline substance has a certain chromatic property) *and* is also an English expression that refers to the state of Oregon. For the same reason *mutatis mutandis*, it is incoherent to stipulate that the English sentence "snow is white" *unambiguously* refers to the property of being true.

1.2.2 To what, if anything, *do* sentences refer? Given that "snow is white" doesn't refer to the property of being true, to what, if anything, *does* it refer? It seems to me—and I'm speaking very tentatively---that, if it refers to anything, that sentence refers to the proposition that snow is white. In virtue of having the form ⌜...snow is white...⌝, a sentence is *ipso facto* about the proposition that snow is white. “If snow is white, then snow isn’t green” says that the proposition that snow is white entails (or otherwise has as a consequence) the proposition that snow isn’t green. "It is not the case that snow is white" says that the proposition that snow is white has the property of being false. "Grass is green and snow is white" says that the proposition that snow is white has the property of being a thing x such that both x and the proposition that grass is green are true. So given the reasonable presumption that E refers O if, in virtue of having the form ⌜...E...⌝, a sentence is to the effect that...O..., it seems to follow that "snow is white" refers to the proposition that snow is white.

And even if it doesn't follow, we'd be doing less damage to the concept of reference by saying that "snow is white" referred to the corresponding proposition than we would be if we said that it referred to its truth-value. Permit me to clarify this statement.

It often happens that, in order for a discipline to move forward, the scope

of a given term must be expanded. To be of use to science, terms such as "temperature" and "acid" must be given meanings that, although consistent with the pretheoretic judgments of laypersons, have a much wider sphere of application than do the concepts associated with those terms by the laity. By defining "temperature" as (I'm simplifying) "mean kinetic energy", the jurisdiction of the former term is increased even though its meaning is otherwise unchanged. Such redefinitions are sometimes referred to as "precisifications" or "delineations."

It's a distinct possibility that, in saying that sentences are referring terms, one is precisifying the term "reference", not identifying its existing meaning. Supposing this true, it's still wrong (or, in any case, bad in some relevant way or other) to say that sentences refer to their truth-values, and it's *less* wrong to say that they refer to the propositions they express. In order to be a precisification of a given term T, as opposed to wholesale redefinition of it, an extension of T's jurisdiction must be consistent (for the most part) with T's existing meaning. (I say "for the most part" since pretheoretic judgments are sometimes revised in light of their own theoretical off-shoots. There are, or could be, objects x such that x is cold according to pretheoretic intuition even though, if temperature is identified with mean kinetic energy, x is extremely hot.) What we know about the *existing* meaning of "refer" is that (the remaining semantic rules of English being what they are) it is in virtue of the fact that "Socrates" refers to Socrates that ⌜Socrates has phi⌝ attributes phi to Socrates; we also know (same qualification) that, if one couldn't attribute phi to Socrates by saying ⌜Socrates has phi⌝, then "Socrates" would not refer to Socrates. In general, we know that to the extent that, in virtue of having the form...E..., a sentence is to the effect that...O..., then E refers to O and that E fails to refer to O to the extent that this condition isn't satisfied. Given this fact, it follows that we'd be horribly *mis*precisifying the

term "reference" were we to say that sentences referred to their truth-values; it also follows that we would not be misprecisifying it *as* horribly were we to say that they referred to the propositions they meant. Give or take some niceties, if S is a sentence that has proposition P for its meaning, then, in virtue of having the form ⌜...S...⌝, a sentence S* is to the effect that...P...This suggests, if only feebly, that, if sentences refer to anything, it is to the propositions they express and, moreover, that it is *not* to their truth-values. [7]

1.3 Davidson on facts Despite everything thus far said, Donald Davidson took it for granted that SS is cogent.[8] And he argued that, given the (alleged) fact that SS is cogent, it follows that there is only one fact. His argument is simple. Put the expression “the fact that” in front of each of (1)-(4). The same principles that show that (1) and (4) must co-refer also shows that “the fact that Mozart was a composer” co-refers with “the fact that Bill Clinton was U.S. President in 1997).”

Instead of rejecting SS, Davidson concluded, absurdly, that these two facts *are* identical and that, by obvious extensions of this reasoning, there is, at most, only one fact. Having, in his mind, established that there is only one fact, he argued that, for this reason, mental states are not representations. Since there is only one fact, Davidson argued, no two mental states represent different facts. On this basis, Davidson concluded that differences between mental states—between, for example, my perception of the piano in my living room and your perception of the ink-marks in front of you—are not representational differences.

Why is this? Because, says Davidson, representations, supposing them to exist, are representations of facts; and, says Davidson, we've seen that there is

only one fact. Therefore, Davidson says, my perception of the piano in my living room doesn't differ in respect of which *fact* it represents from your perception of the ink-marks in front of you. From this Davidson concludes (fallaciously, even relative to his own assumptions), that the one perception is *representationally* just like the other. Thus, the differences between those perceptions –or, by obvious extensions of this reasoning, between *any* two things---have nothing to do with any representational differences that there might be between them. Which means that, if there are any such differences, they are innocuous—in other words, it is not *in terms of* those differences that the palpable differences between those perceptions are to be understood. Which, in its turn, means *either* (i) that neither perception represents anything or (ii) that, if either perception does represent something, it is irrelevant to how that perception differs from the other perception and, therefore, to what sort of thing that perception is. Which, give or take some unimportant nuances, means that neither represents anything.

1.3.1 Evaluating Davidson's contention Davidson's position is obviously false. The fact that I am typing this sentence is distinct from the fact that you are reading it. There is more than one fact. What could be more obvious?

Of course, the word "fact" is a vague one; and, surely, it's at least partly for this reason that Davidson's position has been given any credence. In any case, there is no precisification[9] of the term "fact" that validates Davidson's position.

In some contexts, the word "fact" is synonymous with the word "truth." ("The fact that arithmetic is incomplete was unknown to Frege" is synonymous with "the truth that arithmetic is incomplete was unknown to Frege.") Since there are many truths, this interpretation of the word "fact" doesn't validate Davidson's view.

In other contexts, the word "fact" seems to refer to (re)distributions of mass-energy[10]---to the spatiotemporal conditions and changes in virtue of which certain propositions are true and others are false. Let M be the mass-energy (re)distribution in virtue of which the proposition *x is now asleep* (where x is any sleeping person) is true. Let N be the mass-energy distribution in virtue of which *the Earth is now travelling at approximately 17,000 mph with respect to the Sun* is true. M≠N. That's a just a fact. Pun intended.

Are there are precisifications of the word fact? Probably. (For example, physical laws may qualify as "facts", at least on one interpretation of that word, even though it's not clear that physical laws are identical either with truths or with (re)distributions of mass-energy). But, apart from Davidson's argument to the contrary, there is no reason to expect any of the remaining precisifications to do a better job of validating Davidson's contention than those we've just considered.

1.3.2 Evaluating Davidson's argument for his contention Davidson's argument fails for the same reason *mutatis mutandis* that SS fails. But Davidson’s argument involves a blunder of which SS is innocent. For a mental state to be representational is not for there to be some fact that it represents. Suppose that I hallucinate that there is a pink elephant in my living room. (Assume that, in actuality, there is no such elephant. In other words, suppose my hallucination *not* to coincidentally correspond to reality.) My hallucination is representational. It isn’t like a rock. But there is no elephant x such that my hallucination represents some pink elephant x such that x is in my living room. My hallucination is representational in the sense that it is to the effect that that there exists some pink elephant x such that x is in my living room. For mental state M to be representational isn't

(necessarily) for there to be some fact that M represents; it is (or can be) for M to be to the effect that there exists a fact having such and such characteristics. And the non-existence of such a fact is obviously compatible with the existence of erroneous propositional attitudes to the contrary.

1.4 Neale on a variant of SS In his (ironically named) book *Facing the Facts*[11], Stephen Neale puts forth what he believes to be an improvement on SS.[12] Neale's argument involves the following four assumptions:

G_1: The sentence "Fa" stands for the same fact as the sentence "a=the unique thing x such that (x=a and Fx)." (In this context, "=" is to be taken to mean "is identical with.")
G_2: Any sentence can be put into subject-predicate form.
G_3: What a complex expression refers to is a function of what its parts refer to.
G_4: Definite descriptions are referring terms. "The even prime" is an expression that picks out the number two.

Assume that for some individuals a and b, and some properties G and F, the following are true: "Fa," "Gb," and "a≠b"; and suppose that these sentences stand for facts F_1, F_2, and F_3, respectively.

Bearing these assumptions in mind, consider the following seven sentences.

1. Fa. By assumption.
2. a≠b. By assumption.

3. Gb. By assumption.
4. a=the unique thing x such that (x=a and Fx). Follows from 1, by G_1.
5. a=the unique thing x such that (x=a and x≠b). Follows from 2, by G_2.
6. b=the unique thing x such that (x=b and Gx). Follows from 3, by G_1.
7. b=the unique thing x such that (x=b and x≠a). Follows from 2, by G_1.

1 and 4 stand for the same fact, namely F1. 2, 5, and 7 all stand for the same fact, namely F_2. And 3 and 6 stand for the same fact, namely F_3. "The unique thing x such that (x=a and Fx)" co-refers with "the unique thing x such that (x=a and x≠b)", since both expression refer to a. Therefore, 5 is what results when the definite description in 4 is replaced with a co-referring definite description. Therefore, 4 and 5 refer to the same fact. Since they refer, respectively, to F_1 and F_2, it follows that F_1 and F_2 are the same fact. "The unique thing x such that (x=b and x≠a)" co-refers with "the unique thing x such that (x=b and Gx)," since both expression refer to b. Therefore, 7 is what results when the definite description in 6 is replaced with a co-referring definite description. Therefore, 6 and 7 refer to the same fact. Since they refer, respectively, to F_3 and F_1, it follows that F_3 and F_1 are the same fact. Now, if G_4 is wrong, then definite descriptions are not referring terms, which case 5 is not what results when the definite descriptions in 4 is replaced with a co-referring expression, and 7 is what not results when the definite description in 6 is replaced with a co-referring expression. But assuming that G4 is right, we've established that F_1=F_2. Therefore F_1=F_2=F_3, and "Fa" therefore co-refers with "Gb" and also with "a≠b." So, assuming that definite descriptions are what they appear to be, viz. referring terms, it follows there is only one fact.

As Neale points out, this result, if it is in fact true, is astonishing. But, as Neale then points out, G_2 allows us to go further. Let S and S* be *any* two

distinct statements that are true. (So S and S* needn't have subject-predicate form.) According to G_2, S and be put in subject-predicate form, and so can S*. Given any existing individual a, S can be rendered as "a is an x such that S"; and given any existing individual b, S* can be rendered as "b is an x such that S*." And once that is granted, it follows, by the foregoing argument, that S and S* stand for the same fact. It follows that, given G_1-G4, any two true sentences stand for the same fact.

1.4.1 Evaluating Neale's argument The problems with this argument are very similar to the problems with Davidson's. Consider line 4 of Neale's argument:

4. a=the unique thing x such that (x=a and Fx).

If this sentence is read *attributively* (in other words, if the definite descriptions in them are treated as quantifiers, as opposed to referring terms), then it is logically equivalent with

1. Fa,

in which case those two sentences can therefore be seen as standing for the same fact. But if 4 is read *referentially* (in other words, if the definite descriptions in them are treated as referring terms, as opposed to quantifiers), then 4 stands for the same fact as:

4_R: a=a,

in which case 1 and 4 cannot plausibly be seen as standing for the same fact.

For exactly similar reasons, 5, 6, and 7 must be read attributively if they are can plausibly regarded as standing for the same facts as, respectively, 2, 3, and 4. In general, G_1 is false if 4-7 are read referentially, 4-7 must therefore be read attributively. Bottom line: Neale's argument fails if 4-7 are read referentially. It also fails, for obvious reasons, if they're read attributively.

2.0 Why the Fregean views underlying the Slingshot are incoherent A quick look at an old puzzle will make it clear why SS is unsalvageable. Benjamin Franklin (BF) invented bifocals and he was also the first postmaster general. The semantic rules of English being what they are, it follows that

(CB) "the inventor of bifocals is identical with the first postmaster general."

is correct and that

(TB) "the inventor of bifocals is identical with the inventor of bifocals"

is correct.

But, whereas TB is trivial, CB is non-trivial.[13]

The non-triviality of CB is hard to explain if definite descriptions are supposed to be referring terms. As we saw in Section 1.2, the semantic rules of English being what they are, "the inventor of bifocals" refers to BF exactly if ⌜the inventor of bifocals has phi⌝ attributes phi to BF. So supposing that "the inventor of bifocals" does in fact refer to BF, it follows that CB attributes the property of being identical with the first postmaster general to BF. And supposing that "the first postmaster general" refers to BF, it follows that CB attributes the property of being identical with the

inventor of bifocals to BF. And given both suppositions, it follows that CB says of BF that it has the property of being identical with BF. Which is trivial.

Right now we'll consider Russell's solution to this problem, and we'll then consider Frege's solution. Frege's solution is more easily understood in terms of Russell's than *vice versa*, that being why we're consider Russell's solution first.

2.1 Russell's solution: The Theory of Descriptions According to Russell's Theory of Descriptions (TD) ⌜the inventor of bifocals has phi⌝ says that something uniquely invented bifocals and, moreover, that any such thing has phi. And, according to TD, ⌜ the first postmaster general has phi⌝ says that something was a unique first postmaster general and, moreover, that any such thing has phi. It follows that, if TD is right, CB says the same thing as:

(CB3) “something uniquely invented bifocals, and any such thing is a unique first postmaster general; and something was a unique first postmaster general, and any such thing is a unique inventor of bifocals.”

It is readily seen that CB3 is more succinctly expressed as follows (the “R” stands for “Russell”):

(CBR) “there is some one thing that uniquely invented bifocals and was also a unique first postmaster general.”

To sum up, if TD is right, then CB means the same thing as CBR. It is easily verified that, indeed, CB is true exactly if CBR is true. TD thus seems

to get the right result.[14] Given that CBR is non-trivial—given that historical knowledge is needed to know it—TD readily explains why CB is non-trivial.

2.2 Frege's solution

Here is Frege's solution to this problem:

(FR)[15] **"The inventors of bifocals" is a referring term; it picks something out. That expression refers to a given thing if and only if that thing has the property of being a unique bifocal-inventor. If a given thing didn't invent bifocals then "the inventors of bifocals" doesn't refer to it. If lots of different individuals invented bifocals, then "the inventors of bifocals" doesn't refer to anything. (If ten people collectively wrote *War and Peace*, it would be wrong to speak of *the* author of that book.) And if a given thing *did* uniquely invent bifocals, then "the inventors of bifocals" *does* refer to it.**

So there are *two* dimensions to the semantics of "the inventors of bifocals." There is what it refers to: this is the referent of "the inventors of bifocals." (Sometimes "reference" is used instead of "referent.") And there's the property that a thing must have to *be* its referent; this property is the *sense* of that expression. The referent is BF, and the sense is the property of being a unique bifocal inventor.

Analogues of these points hold with respect to "the first postmaster general." There is its referent (which is BF); and there is its *sense* (which is the property of being a unique first postmaster general).

The reason that TB is non-trivial is that, although the definite

descriptions in it have the same referent, they have different senses. A comparison may help. Imagine that you're looking at a certain person in the distance. That person is in fact Ralph, but you don't know it, because he's so far away and also, let us suppose, because he is (for reasons that aren't important) wearing a suit of armor. You see that he is doing jumping jacks. Later Ralph is standing right in front of you and, of course, you see him. He's wearing a plaid shirt. Once again, he's doing jumping jacks.

The first visual perception has the same object as the second. And each of those perceptions says of that object that he or she is doing jumping jacks. But those two perceptions pick out that object in different ways. In the first case, that object is presented to you as something that has one set of properties (it's far away, it's wearing a suit of armor, etc.); in the other case it's presented as having a different set of properties (it's nearby, it's wearing a plaid shirt, etc.). And that's why, even though those perceptions have the same object, and even though both attribute the property of doing jumping jacks to that object, those perceptions tell you different things.

Something similar holds of "the inventors of bifocals" and "the first postmaster general." They have the same object (referent); but they pick that thing out through different bodies of information. That's why ⌜the inventors of bifocals has phi⌝ and ⌜the first postmaster general has phi⌝ tell you different things, *even though* they attribute the same property (phi) to the same object (BF).[16]

2.2.1 Evaluating Frege's solution This all seems not only plausible, but practically self-evident. But, so far as what Frege says doesn't coincide with

TD, what Frege is saying is false and, indeed, incoherent.

Let's consider the two sense-perceptions discussed in FR. The one perception tells you that there is something in the distance, that it's wearing a suite of armor, etc., and that, moreover, that thing is doing jumping jacks. So the sense in which Ralph is the *object* of that perception is that Ralph in your perception *describes* Ralph. Your perception says: "something or other is over there, in the distance, wearing a suit of armor, [etc.]." Who (or what) fits that description? In other words, who has the relevant properties? Ralph. So Ralph is the object of that perception in the sense it *describes* him.

The information borne by that perception has the form "something has such and such properties [it's wearing a suit of armor etc.], and any such thing is doing jumping jacks." Thus, the content of that perception is given by an existence-claim (a claim of the form "something has phi"). Ralph uniquely satisfies that existence-claim. (In other words, he and he alone has the relevant property.) And it is in virtue of his doing so that he is the object of the perception in question.

The same thing *mutatis mutandis* is true of your second perception. The information borne by that perception is along the lines of "something has thus and such properties [it's standing nearby, is wearing a plaid shirt etc.], and any such thing is doing jumping jacks." Ralph is the object of that perception by virtue of the fact that he and he alone has the relevant properties—that he and he alone is a thing standing in the relevant place while wearing a plaid shirt, etc.

In general, when a given thing x is picked out *by way of* descriptive information, what's really going on is that some statement of the form "something has such and such properties" is being affirmed (not necessarily verbally), and x is what fits the description.

If we bear this point in mind, we find that, so far as it isn't totally

incoherent, Frege's analysis collapses into Russell's. Consider the sentence:

(IB[17]) "the inventor of bifocals is tall."

By Frege's own admission, "the inventors of bifocals" picks a given thing out if and only if that thing fits the relevant description (i.e., if and only if that thing is a unique inventor of bifocals). And if there is a unique inventor of bifocals, that thing, whatever it is, is the thing to which "the inventors of bifocals" refers. So supposing that there is such a thing and that x is it, IB describes x as being tall.

It goes without saying that, if there is no unique inventor of bifocals, then there is no way that IB can be true. Frege himself emphasizes this last point. (And he's right to do so. There's no way that "the elephant in the corner is eating peanuts" can be true, if there's no elephant in the corner.)

The upshot is that IB is true if and only if:

(IB*) there is a unique inventor of bifocals and, moreover, that thing, whatever it is, is tall.

But what does TD say about IB? According to TD, the meaning of IB is: *something uniquely invented bifocals, and any such thing is tall.* Which is precisely what FB says. So Frege's theory collapses into Russell's.

By obvious extensions of this, it follows that, if Frege's theory is right, then CB's meaning is identical with that of CBR. So Frege's theory accounts for the non-triviality of CB in the same way as Russell's theory (as TD).

2.2.2 Evaluating Frege's solution (continued) For argument's sake, suppose that the prettiest living woman is identical with most evil living person.

Given only that

(1) “Smith wants to marry the prettiest woman on the planet”

is true, it doesn't follow that:

(2) “Smith wants to marry the most evil person on the planet”

is true. But why not? Because what (1) really says is that:

(1*) Smith wants it to be the case that Smith is married to a woman who is prettier than anyone else,

and what (2) really says is that

(2*) Smith wants it to be the case that Smith is married to somebody who is more evil than anyone else.

(1*) makes it clear that, if (1) is true, the object of Smith’s desire is not a *person* and is instead a *proposition*. And (2*) makes it clear that, if (2) is correct, Smith wants a certain proposition to hold, namely, *that Smith is (or will be) married to a woman who is prettier than anyone else*. And (2*), if correct, makes it clear that the object of Smith’s desire is not a *person*, but a proposition, namely, *that Smith is married to somebody who is more evil than anyone else*. Since those two propositions *are* different, the fact that one of them has the property of being desired to be true by Smith, whereas the other does not lack that property, does *not* constitute a violation of Leibniz’s Law (*if x=y, x has a given property iff y has that property*).

Few authors, if any, deny that apparent exceptions to Leibniz's Law are to be solved in this way. So given that, according to TD, the meanings[18] of (1) and (2) are given by (1*) and (2*), and given that, nuances, aside, there's no other way to deal with the datum that (1) and (2) can simultaneously be true, it's odd that authors have continued to countenance Frege's system.

2.2.3 Special compositionality in relation to Frege's theory There are other, related problems with that system According to the principle of special compositionality (SC), what a referring term picks out is a function of what its parts pick out.[19] Bearing this in mind, let's suppose, as before, that the prettiest woman in existence is also the most evil person in existence. And let's also suppose, with Frege, that definite descriptions are referring terms. Given these suppositions, it would seem to follow that the underlined part of:

(3) Smith believes that the prettiest woman in existence is identical with the prettiest woman in existence

refers to the same proposition as

(4) Smith believes that the prettiest woman in existence is identical with the most evil person in existence.

But the underlined part of (3) very much seems *not* to co-refer with the corresponding part of (4).

There are only three ways to go here. (i) Deny SC. (ii) Deny that the underlined parts of 3 and 4 refer to different propositions. (I.e. take the view

that they co-refer.) (iii) Deny that 4 is what results when a referring term in 1 is replaced with a co-referring term.

Understandably, Frege rejects each of (i) and (ii) and thus accepts (iii). (To this extent, his semantics coincides with Russell's, give or take some technicalities relating to the fact that, according to TD, each of 3 and 4 is syntactically ambiguous.) Frege thus says that the definite descriptions occurring in 3 and 4 refer, not to some person, but rather to their "customary senses." What he means is that, in each of 3 and 4, "the prettiest woman in existence" refers, not to Sally or Betty or Ethel, but rather to the property (or concept) of being a woman who is prettier than every other (or some such[20]); and, in 4, "the most evil woman in existence" refers to the property (concept) of being a person who is more evil than any other.

An immediate consequence of Frege's theory is that the word "snow" has three utterly distinct meanings in:

(5) "snow is white and Jerry believes that snow is white and Fred believes that Jerry believes that snow is white."

The first occurrence of "snow" in 6 refers to snow. The second to the sense of the first occurrence. The third refers to the sense of the second occurrence.

But surely all three occurrences of "snow" in (6) mean the same thing. Frege's is clearly jumping through hoops to make his bankrupt theory work.

Another immediate consequence of Frege's theory is that the word "snow" doesn't refer to snow in:

(6) it is true that snow is white.

But that's obviously false. (In any case, there's no independent support

for this claim.[21]) The occurrence of “snow” in (6) clearly has the same meaning as its counterpart in “snow is white.”

In any case, what does it even mean to say that, in (5), the third occurrence of “snow” *refers* to the sense of the sense of “snow.” If you say “Socrates was bald,” you are attributing baldness to Socrates—not to Aristotle or Plato or anyone else who isn’t Socrates. If you say “Socrates was intelligent,” you are attributing intelligence to Socrates and Socrates alone. Thus, “Socrates” refers to Socrates because, the remaining semantic rules of English being what they are, ⌜Socrates has psi⌝ attributes psi to Socrates. By parity of reasoning, if the third occurrence of “snow” in (5) really referred to a second-order sense, then it would make some statement about that sense—it would attribute a property to it. But it doesn’t. It doesn’t say anything about such a sense. Whatever it is that it does in connection with such a sense, it isn’t comparable to what “Socrates was intelligent” does in connection with Socrates.

Finally, to the extent that Frege's views concerning indirect reference are intelligible, they undermine his own semantics. According to Frege, the proposition meant by

(#) "the prettiest woman in existence is smart"

is true just in case:

(##) somebody who is smart is uniquely a woman who is prettier than any other.

In general, the proposition meant by:

(*) ⌜...the phi... ⌝

is true just in case:

(**) something x such that...x...uniquely has phi.

Let "FR" (short for "Frege-Russell") be the principle ("hypothesis" might be a better word) that the meaning of (*) is given by some proposition that is equivalent (if not identical with) (**) as FR. It is to be emphasized that Russell and Frege *both* accept FR.

Here's where they differ. According to Russell's system (TD), ⌜the phi ⌝ is in the same category as "something", "everyone", etc. According to Frege's system, ⌜the phi ⌝ is *not* in that category.

Here's the problem. Let O be an occurrence of "the prettiest woman in existence" that falls within the scope of four epistemic (or modal) operators. So O might be the underlined expression in:

(&) "Smith believes that Sally knows that Larry hopes that Mary is certain that the prettiest woman in existence is smart."

A consequence of TD *and* of Frege's system is that the semantics of the occurrence in (&) "the prettiest woman in existence" is given by FR. But if Frege's system is correct, then that occurrence refers, not to Sally (who, we'll assume, is the prettiest woman in existence), but to the "customary sense" of the customary sense of the customary sense of the customary sense of the occurrence in (#) of "the prettiest woman.” (Let O# be the just-mentioned occurrence.) So, if Frege's system is right, then the semantics of each of O and O# is given by FR *even though*, whereas the one occurrence

refers to Sally, the other refers to a fifth order property (or concept). Thus, a consequence of Frege's theory is that, in respect of their semantics, two expression-occurrences can be functioning in the very same way *even though* the one occurrence refers to an individual while the other refers to an nth order property (for any whole number n). So a consequence of Frege's system is that there can be reference-shifts where there haven't been semantic shifts. Which is obviously incompatible with the principle (thought by many authorities, e.g Dummett (1973), to be (part of) the meaning of Frege's celebrated "context principle") that, it is in virtue of how a given expression E affects the truth-conditions (or meaning) of sentences in which it occurs that E refers to this as opposed to that object. In Section 1.2, we saw reason to believe that for "Socrates" to refer Socrates is for ⌜....Socrates...⌝ to be to the effect that ...Socrates...---that this condition's being satisfied is necessary and sufficient for that expression's referring to that individual.

Basically, for a given expression to have a given referent is for it to have certain effects on the truth-conditions of the sentences in which it occurs. A corollary is that there, where those changes haven't occurred, there hasn't been a shift in reference, a related corollary being that, where they have occurred, there has been such a shift. For any property phi, and for any whole number n, the semantics of an occurrence O^ that falls within the scope of n epistemic (or modal or otherwise intensional) operators of the phi is given by FR. Thus, so far as Frege's clearly contrived and *ad hoc* theory has any intelligible meaning, it's clearly false.

2.3 A final note concerning SS SS assumes that:

(i) each of the two definite descriptions occurring in it is a singular term.

SS makes the additional assumption that

(ii) reference isn't changed by intersubstituting co-referring terms that *have different senses*.

And, of course, SS assumes that

(iii) referring terms can have both sense and reference (it assumes that the definite desriptions occurring in the second and third sentences have both sense and reference).

(i) is innocuous. Even though, for reasons already discussed, some hold that definite descriptions are *not* singular terms, being instead quantifiers, we could obviously *stipulate* that the definite descriptions in SS are referring terms. We could stipulate that, for any property phi, and for any object O, ⌜the phi⌝ refers to O exactly if O uniquely has phi.

But this wouldn't be enough to get SS go through. For argument's sake, suppose that ⌜the phi⌝ refers to O exactly if O uniquely has phi; and *further* suppose that, in virtue of having the form ⌜…the phi…⌝, a sentence is true exactly if…O…In that case, supposing that O were a unique phi, ⌜Plato thought highly of the phi⌝ would bear a proposition that was true exactly if Plato thought highly of O, it being entirely irrelevant what properties (other than being thought highly be Plato) O had, and ⌜the phi was a great wrestler⌝ would bear a proposition that was true exactly if O was a great wrestler, it being entirely irrelevant what properties (other than being a great wrestler) O had.

Thus, supposing that ⌜...the phi…⌝ were defined in the way thus proposed, there is some x such that x=Socrates such that:

(*) “Plato thought highly of the author of the great philosopher who died of hemlock poisoning"

would be true iff:

(**) Plato thought highly of x.

SS therefore assumes, not just that ⌜the phi⌝ is a singular term, but *also* that that, in virtue of having the form ⌜…the phi…⌝, a sentence be true just in case: exactly one thing O has phi and…x…

But, as we just saw, that assumption, which is the essence of Frege's semantics, is not a coherent one; and we may conclude, if only tentatively, that SS is a spurious attempt to establish a conclusion for which (see Section 1.2) there is no independent evidence and for whose negation there *is* independent evidence (*op. cit.*).

Bibliography

The Philosophy of Language. (1990.) Second Edition A.P. Martinich (ed.). Oxford: Oxford University Press.

Anderson, C. Anthony. 1998. “Alonzo Church’s Contributions to Philosophy and Intensional Logic”. *Bulletin of Symbolic Logic. Volume 4. Number 2. June 1998.*

Barwise, Jon and Perry, John.1983. “Semantic Innocence and

Uncompromising Situations". In *The Philosophy of Language.* (1990.) Second Edition A.P. Martinich (ed.).
Oxford: Oxford University Press.

Beaney, Michael (ed.). 1997. In *The Frege Reader.* Oxford: Blackwell.

Church, Alonzo. "Carnap's Introduction to Semantics". *The Philosophical Review.* Vol. LII. No. 3. 1943. 298-304.
Church, Alonzo. 1956. *Introduction to Mathematical Logic.* Princeton: Princeton University Press.

Davidson, Donald. 1984. "True to the Facts". In *Essays on Truth and Interpretation.* Oxford: Oxford University Press.
Davidson, Donald. 1967. Truth and Meaning. *Synthese* 17, 304-323.
Davidson, Donald. 1980. *Inquiries into Truth and Interpretation*. New York: Oxford University Press.
Frege, Gottlob. 1892a. On *Sinn* and *Bedeutung.* In Michael Beaney (Ed.), *The Frege Reader*, (pp. 151-171). Oxford: Blackwell.
Meaning and Truth. 1991. Jay L. Garfield and Murray Kitely (eds.) St. Paul, Minnesota: Paragon House

Dummett, Michael. 1973. Frege: Philosophy of Language. Cambridge MA: Harvard.

Gödel, Kurt. 1944. "Russell's Mathematical Logic." In *The Philosophy of Bertrand Russell.* Edited by Paul Arthur Schilpp. Evanston, Illinois: Northwestern University Press.

Neale, Stephen. 2001. *Facing the Facts.* Oxford: Clarendon Press.

Perry, John. 2000. *The Problem of the Essential Indexical and Other Essays.* Palo Alto: CSLI Publications.

Quine, Willard van Orman. 1953. “Reference and Modality.” In *From a Logical Point of View.* Cambridge: Harvard University Press.
--1960. *Word and Object.* Cambridge: The M.I.T. Press.
-- 1966. “Three Grades of Modal Involvement.” In *The Ways of Paradox.* Cambridge: Harvard University Press.

Russell, Bertrand.1905. “On Denoting”. In A.P. Martinich (ed) *The Philosophy of Language* Oxford: Oxford University Press.

Strawson, Peter. 1950. ‘On Referring’. In A.P. Martinich (ed.) *The Philosophy of Language.* Oxford: Oxford University Press.

Taylor, Kenneth. 1998. *Truth & Meaning.* Oxford: Blackwell.

Implicit Comparatives and the Sorites

Abstract. A person with one dollar is poor. If a person with n dollars is poor, then so is a person with n+1 dollars. Therefore a person with a billion dollars is poor. True premises, valid reasoning, a false a conclusion: a paradox. This is an instance of the Sorites-paradox. Most attempts to solve this paradox

reject some law of classical logic, usually the law of bivalence. I show that this paradox can be solved while holding onto all the laws of classical logic. Given any predicate that generates a Sorites-paradox, significant use of that predicate is actually elliptical for a relational statement: a significant token of 'Bob is poor' means *Bob is poor compared to x*, for some value of x. Once a value of x is supplied, a definite cut-off line between having and not having the paradox-generating predicate is supplied. This neutralizes the inductive step in the associated Sorites argument, and the would-be paradox is avoided.

I. A person with an IQ of 70 is not smart. For any number n, if a person with an IQ of n is not smart, then a person with an IQ of n+1 is not smart. Thus a person with an IQ of 455 is not smart. But this conclusion is false. So we have a false conclusion following from true premises.

A person with one dollar is not wealthy. For any number n, if a person with n dollars is not wealthy, then a person with n+1 dollars is not wealthy. Thus a person with 777 million dollars is not wealthy. True premises, valid reasoning, false conclusion: a paradox.

These are examples of the Sorites paradox. This essay will set forth a solution to this paradox. This solution is not in the literature and is not derivative of anything in the literature.[22]

A disclaimer is appropriate. There are many different Sorites-type paradoxes – we've already considered two different ones. I do not know whether what we will say constitutes a solution to *every* paradox belonging to this class. But I am fairly certain that it constitutes a solution to some of them.

The Sorites paradox is an ancient one. It was first put forth by a contemporary of Aristotle's, one Eubilides of Miletus. But no serious attempts at a *solution* were made until the last thirty years or so.[23] Almost all of the solutions that have been offered involve the abandonment of some accepted law of logic. According to one of these solutions, the existence of Sorites-paradoxes show that some propositions are neither true nor false. According to another solution, such paradoxes show that propositions admit of *degrees* of truth: some propositions are 70% true, others are 98% true, and so on.[24]

Both of these solutions involve the abandonment of the law of excluded middle: any proposition is either 100% true or 100% false. Of course, this principle is one of the cornerstones of classical logic.

Presumably the laws of classical logic should not be rejected if there is any way to hold onto them. My solution to the Sorites does *not* involve any abandonment of those laws and is therefore preferable (*ceteris paribus*) to a solution that abandons them.

Sorites arguments have the following structure. They consist of (i) a 'base' clause ('someone with an IQ of 70 is not smart'), (ii) an 'inductive clause' ('if a person with an IQ of n is not smart, then a person with an IQ of n+1 is not smart'), and a conclusion that follows from the premises but is plainly false ('a person with an IQ of 455 is not smart').

The argument set forth in this paper is basically this. The base clause of any Sorites argument is implicitly a *comparison* of some kind: it involves comparison to some standard or to some other entity. Once it is made explicit what the standard or object of comparison is, the inductive clause is immediately seen to be false: for once that standard is identified, a definite cut-off line between having and not having the offending property is thereby

established.

Consider the base-clause in the argument set forth in the first paragraph of this paper. It is meaningful to ask: ‘a person with an IQ of 70 is not smart *compared to what*?’ And it would be meaningless to respond by saying: ‘not compared to anything; such a person is not smart *simpliciter*.’ No—a person with an IQ of 70 is very stupid compared to the average person but astronomically and incalculably intelligent compared to the average potato bug. So any significant utterance of ‘a person with an IQ of 70 is not smart’ involves reference to some standard; it is elliptical for some proposition like ‘a person with an IQ of 70 is not smart *compared to the average person*.’ Of course, once that italicized qualification is added to the base-clause, corresponding qualifications must be added to the other two clauses, lest the argument involve an equivocation. So the inductive clause and the conclusion become: (ii) ‘if a person with an IQ of n is not smart *compared to the average person*, then a person with an IQ of n+1 is not smart *compared to the average person*’ and (iii) ‘a person with an IQ of 455 is not smart *compared to the average person.*’

But notice that (ii) is simply false. Suppose that the average IQ of human beings is 102. In that case, there is some n such that a person with an IQ of not smart compared to the average person, and such that a person with an IQ of n+1 *is* smart compared to the average person. It is true for n=101.5, n=101.6, and so on. So the inductive clause is simply false. And the paradox is thereby avoided.[25]

I submit that a corresponding treatment of many other Sorites arguments will show each those arguments to contain a false premise (the inductive clause). Thus the associated paradox vanishes; for it is in no way paradoxical that a false conclusion should follow from false premises. The purpose of this essay is to give a rigorous and comparatively general statement of the

argument just given.

II. Some preliminary remarks are in order. Not all properties generate Sorites-type arguments. Only properties that can be had to varying degrees generate them. The property of weighing 323 lbs is binary: one either has it or one doesn't. One can't weigh 323 lbs to a greater extent than another person. 'Binary' properties --- those that don't admit of comparatives – don't generate Sorites arguments. But, as far as I can tell, all properties that do admit of comparatives generate them: fatness, slowness, dirtiness, smartness.

Given any statement of the form 'x is P', where P is a property of the kind that generates a Sorites paradox, let us describe that sentence as a 'Sorites sentence'. Examples of such sentences are 'that woman is smart', 'that man is tall', 'that object is moving quickly.' Further, we will describe properties like P as 'Sorites-properties'; and we will describe predicates like 'P' (the expression denoting P) as 'Sorites-predicates.'

Given a Sorites-sentence, the sentence formed by negating the predicate is itself a Sorites-sentence. 'Bob is wealthy' is a Sorites sentence. So, therefore, is 'Bob is not wealthy'.[26] And prefixing a 'not' onto any Sorites-type predicate results in a Sorites predicate. 'Smart' is a sorites-type predicate; so, therefore, is 'not smart.' (Thus, the property of being not smart is a Sorites-property.)

III. Let us now proceed to the argument proper. Sorites-sentences are context-sensitive. The sentence 'Bob is fast' makes a false statement if Bob is a cheetah whose top running speed is 30 mph. That same sentence makes a true statement if Bob is a turtle whose top crawling speed is 2 mph. In a very poor country, the sentence 'Bob is not wealthy' may be false if Bob has $100 to his name. In a very rich country, the sentence 'Bob is not wealthy' may be

true if Bob has $90,000 to his name.

This context-sensitivity is easily explained when it is realized Bob cannot be wealthy *simpliciter*. He must be wealthy compared to most Albanians, or wealthy compared to most Americans, or wealthy enough to take adequate care of his children (who have medical conditions requiring expensive treatment). It would be meaningless to say that Bob was *just* wealthy. Pick any amount of money. For any amount of money M (other than no money at all), there will be some context such that a person with M can be truly characterized as 'wealthy' and a different one where a person with M can be truly characterized as 'not wealthy.'

What this tells us is that 'wealthy' (and therefore 'not wealthy') is not really a one-place, but a two-place, predicate: 'x is wealthy relative to some standard', 'x is wealthy compared to y.'

This applies mutatis mutandis to any Sorites sentence. Suppose you say 'Bob is smart.' I can legitimately ask: 'Smart compared to what? Smart compared to most PhD's in physics? Smart compared to most people?' Given any sentence of the form 'x is smart', there are some contexts where that would constitute a true statement and some where it would not. (This is the case unless x has absolutely no intelligence at all. We will deal with this problem below). This tells us, I submit, that sentences of the form 'x is smart' really express propositions that relativize x's intelligence to some standard: x's intelligence is being compared to that of other PhD's in physics, to other people in the remedial reading class, to other dogs, or other iguanas.

So different tokens of 'x of smart' involve comparison to different standards, and thus express what are, strictly speaking, different propositions. If Bob is a dog, and you say 'Bob is smart' you are probably saying 'Bob is smart *for a dog*'. If you say 'Bob is smart' and Bob is a Nobel laureate, you are probably saying 'Bob is smart for a person'. In any case, if Bob is a

Nobel Laureate, and you say ‘Bob is smart’, you are probably not saying that Bob is smart *compared to a dog.* That is why the truth-conditions for ‘Bob [a dog] is smart’ are not the same as ‘Bob [the Nobel Laureate] is smart’. They have different truth-conditions because they express different propositions: *Bob is smart compared to the average dog* and *Bob is smart compared to the average human.* This point applies to many, if not all, sentence of the form ‘x has phi’, where ‘phi’ is a Sorites-predicate.

Let us apply these considerations to some specific Sorites-argument:

(i) An object moving at one mph is not moving quickly. (ii) If an object moving n mph is not moving quickly, then an object moving n+1 mph is not moving quickly. (iii) So an object moving 400 million mph is not moving quickly.

Let us refer to this argument as S.

The first sentence of S, if used significantly, must (elliptically) express a proposition that *relativizes* the speed of an object moving one mph to some standard – that compares the speed of such an object with that of some other object (or class of objects). For it is legitimate to ask: *Relative to what* is an object moving at one mph not moving quickly? Compared to a glacier? No. Compared to a light-ray? Yes.

So someone who was putting forth S would have say *relative to what* an object moving one mph was not moving quickly. So, in effect, the base clause of S would be some comparative statement like:

(i’) An object moving one mph is moving slowly compared to the average horse.

The validity of an argument depends on the terms in it being used with a uniform meaning. If ‘quickly’ in (i) is elliptical for *quickly compared to the average horse*, while ‘quickly’ in (ii) is elliptical for *quickly compared to the average meteor*, then S cannot possibly have any validity. So ‘quickly’ in each of (i), (ii) and (iii) must be used with the same meaning – it must be elliptical for the *same* thing in each case. Consequently, we must make the appropriate substitutions. So S then becomes (say):

(i’) An object moving at one mph is not moving quickly compared to the average horse (ii’) If an object moving n mph is not moving quickly compared to the average horse, then an object moving n+1 mph is not moving quickly compared to the average horse. (iii’) So an object moving 400 million mph is not moving quickly compared to the average horse.

But once these substitutions are made, it becomes clear that (ii’) – the inductive clause – is simply false. For *there is* some number n such that, if an object x is moving at n mph, then x is *not* moving faster than the average horse, and such that if x is moving n+1 mph, then x *is* moving faster than the average horse. Suppose that the average horse has a maximum speed of 45 mph. In that case, if x is moving 44.5 mph, then x is not moving faster than the average horse, and if x is moving 45.5 mph, then x *is* moving faster than the average horse.

It is pretty clear that these considerations apply *mutatis mutandis* to any Sorites predicate. Given any use of a predicate that admits of comparative – and all Sorites predicates are of this kind – it is legitimate to ask for a standard of comparison. If you say ‘that rock is heavy’, I can legitimately

ask: ‘heavy compared to what?’ In other words, I can ask for a standard of comparison. And once a standard is supplied, then the associated Sorites-generating inductive clause is blocked. ‘For all n, if an object weighing n lbs is not heavy *compared to the average human*, then an object weighing n+1 lbs not wealthy *compared to the average human*.’ But that statement is false. If the average American weighs 145 lbs, then an object weighing 144.5 lbs will not be heavy compared to the average human while an object weighing 145.5 will be heavy compared to the average human.’

IV. Let us now consider some objections to our argument. The major resistance to it will, I think, be this (single-spaced statements are those of an imaginary objector):

Despite what you’ve said, I think that sentences like ‘Bob is tall’, ‘that rock is heavy’, and ‘Fred is wealthy’ – Sorites-type sentences – have a non-comparative meaning. It *is* meaningful to say ‘Bob is wealthy *simpliciter*.’ The same applies to all other sentences involving Sorites-type predicates

I’ve already given one argument against this. Here is another. Consider the following two sentences:

(A) Bob is tall.

(B) Bob is taller than Fred.

(B) is synonymous with

(B’) Bob is more tall than Fred.

And (B) is also synonymous with

(B’’) Bob is tall *compared* to Fred.

In (B’’) ‘tall’ is really a two-place predicate. In other words, in (B’’), the word ‘tall’ must be thought of as a predicate with two argument places. Now surely the word ‘tall’ has the same meaning in (A) that it has in (B’’). There is obviously no ambiguity at work here. (It is not like the word ‘bank’, which has two different meanings.) Now if ‘tall’ were a two-place predicate in (B’’), and a one-place predicate in (A), then it couldn’t have the *same* meaning in both (A) and (B’’).

Let me clarify this. Obviously ‘tall’ is functioning in a completely orthodox way in (B’’). And obviously ‘tall’ is a two-place predicate in (B’’). Finally, ‘tall’ clearly has the same meaning in both (A) and (B’’). Given these facts, it is hard to see how ‘tall’ could not be functioning as a two-place predicate in (A). This shows that ‘tall’ is really elliptical for ‘tall *compared to*__’.

As against this, one might reason as follows:

It is true both that ‘tall’ has the same meaning in both (A) and (B’’). It is also true that ‘tall’ is part of a two-place predicate in (B’’). Nonetheless ‘tall’, in (A), is a one-place predicate, and is not part of a two-place predicate. What is going on is that the aforementioned two-place predicate in (B’’) is being constructed out of the one place predicate in (A).

The problem with this proposal is not hard to see. The meaning of a

predicate is intimately bound up with the number of argument places it has. [27] If P1 and P2 are predicates of different degrees (i.e. if they have a different number of argument places), then *ipso facto* they have different meanings. And I don't see how the obviously two-place predicate '__ is tall compared to__' could be constructed out of a one place predicate.

The natural resolution of this is the one we considered a moment ago. The term 'tall' has the same meaning in both (A) and (B'). And this predicate doesn't somehow retain the same meaning while shifting the number of argument places it has, or while somehow being assimilated into a predicate of a different degree than it originally had. What we say is that 'tall' is always comparative: x is tall *compared* to some standard. Sometimes the comparison is explicit. Sometimes it is implicit. Presumably, it is left implicit when the standard can be taken for granted, which is often the case.

Admittedly, there are apparent problems with the line of reasoning just set forth:

Consider the predicate 'acidic.' Some things can be more acidic than others. So statements of the form 'x is more acidic than y' are significant. But it is obviously meaningful to say 'x is acidic' *simpliciter.* At the same time, 'acidic' obviously has the same meaning in 'x is acidic' – where 'acidic' is functioning as a one-place predicate – that it has 'x is more acidic than y' – where 'acidic' is part of a two-place predicate. So there must be some error in your reasoning. Exactly similar remarks apply to 'electrified', 'malleable', 'ductile' and various other physical terms. Also, it seems to apply to the term 'intelligent'. We speak of a life-form as being 'intelligent' *simpliciter*, meaning that it has *some* intelligence. But we can also meaningfully speak of one thing as being *more* intelligent than another.

When it is said that 'x is acidic', what is really meant is: 'x is acidic to

some positive degree' or 'for some positive number n, x is acidic to degree n'. Degree-statements (even highly generic ones, like 'x has P to *some* positive degree or other') are really statements about the *measure* of something. And relativization to a standard is the essence of measurement: to measure is to compare – to compare with something taken as a standard. So statements about measurements are comparative statements. Given this, it follows that degree-statements – even highly generic ones like 'x is acidic *to some positive degree or other'* -- always presuppose some benchmark (in a quite literal sense of 'benchmark', in some cases). So to say 'x is acidic' *is* in fact to make a kind of comparative statement.

I'd like to consider another *prima facie* problem with our analysis. If somebody has *no* money, then it certainly seems reasonable to say that she is non-wealthy *in an absolute sense.* The same is true *mutatis mutandis* of all other Sorites-type predicates. And this seems to controvert my point that 'wealthy', 'non-wealthy', 'tall', 'smart', and so on, have only a *comparative* meaning.

This point does not go through. We've seen reason to believe that 'non-wealthy' is typically used to make a comparative statement. So if we say that 'non-wealthy' has a truly *absolute* and non-comparative meaning, then we are really saying that it has *two* meanings: we are saying that 'non-wealthy' is sometimes a non-comparative predicate and is sometimes a comparative predicate, and is thus ambiguous between two very different significances.

But it seems extremely arbitrary to say that 'non-wealthy' is sometimes an absolute (one-place) predicate and sometimes a comparative (two-place) predicate. Obviously 'non-wealthy' has the same meaning in 'Bob is *absolutely* non-wealthy' (meaning: he has *no* money) that it has in 'Bob is comparatively non-wealthy' (meaning: he has less money than others in a certain reference class). And there is no reason to believe that the term 'non-

wealthy' corresponds to predicates with different numbers of argument places in these two sentences; it is simply not natural to suppose that such a dramatic shift in its semantics occurs.

The natural way to go is this. It is in some sense true to say that a person who has no money is non-wealthy in an 'absolute sense'. But to say that Bob is non-wealthy in an absolute sense is not to say that he is non-wealthy in some way that involves reference to *no* standard. On the contrary, it is to say that he is non-wealthy relative to *every* standard. So being 'absolutely' non-wealthy involves just as much a reference to some standard of comparison as being comparatively non-wealthy.

I would like to end this essay by discharging an assumption that I have relied on throughout.[28] I have said that

(i) 'x is smarter than y'

is equivalent to

(ii) 'x is smart *compared to* y'

and also to

(iii) 'x is smart *relative to* y'.

This assumption is obviously crucial to my argument. Obviously 'smarter' is a two-place predicate. (It would be meaningless to say 'Sam is smarter', unless that were elliptical for 'Sam is smarter than Bob' or some other relational sentence.) It is essential to my argument that 'smart' is really elliptical for some *two*-place predicate that has the same logical form as

‘smarter’. My judgment is that ‘__is smart *compared to*__’ and ‘__is smart *relative to*__’ are such two-place predicates. But this judgment might be questioned, for one might say:

Offhand, it does *not* seem to me that the statements ‘x is smarter than y’ has quite the same meaning as ‘x is smarter compared to y’ or ‘x is smart relative to y’.

My response is this. Perhaps[29] it may legitimately be questioned whether ‘x is smarter than y’ has the same meaning as ‘x is smart compared to y’ or ‘x is smart relative to y’. But that isn’t ultimately relevant. The important point is that one obviously *could* construct a two-place predicate out of ‘smart’ that had the same meaning as ‘smarter’.[30] Once it is granted that ‘smart’ is elliptical for *some* such two-place predicate, then my argument in its entirety goes through. Now it may be that ‘__is smart *compared to*__’ is not that predicate. But as long as *some* two-place predicate can be constructed out of ‘smart’ that has the same meaning as ‘smarter’, then my argument goes through. (Whether the English language *already* contains such a predicate is irrelevant.) Now clearly there is no barrier to *creating* some predicate out of ‘smart’ that has the same logical form as ‘smarter’: let ‘__is smart **kompared to**__’ be that predicate. (So ‘x is smart **kompared to** y’ has the same logical form as ‘x is smarter than y’.) Given the absence of such a barrier, it is plain that ‘smart’ *is* elliptical for a two-place comparative predicate. (For as we, in effect, saw earlier, it wouldn’t be possible to *construct* a two-place predicate out of ‘smart’, unless ‘smart’ were *already* such a predicate.) And, as we’ve seen, once this is granted, then our analysis goes through in its entirety.

References

Cargile, James. 1999, ‘The Sorites Paradox’. in *Vagueness: A Reader*, Rosanna Keefe and Peter Smith (eds.), Cambridge: The M.I.T. Press.

Dummett, Michael. 1978, ‘Wang’s Paradox’, in *Truth and Other Enigmas*, Cambridge: Harvard University Press.

Fine, Kit. 1999, ‘Vagueness, Truth and Logic’, in *Vagueness: A Reader*, Rosanna Keefe and Peter Smith (eds.), Cambridge: The M.I.T. Press.

Parsons, Terence. 2000, *Indeterminate Identity: Metaphysics and Semantics*, Cambridge: The M.I.T. Press.

Sorensen, Roy. 1998, *Blindspots*, Oxford: Clarendon Press.
--2001, *Vagueness and Contradiction*, Oxford: Oxford University Press.

Tye, Michael. 1999, ‘Sorites Paradoxes and the Semantics of Vagueness’, in *Vagueness: A Reader*, Rosanna Keefe and Peter Smith (eds.), Cambridge: The M.I.T. Press.

Williamson, Timothy. 1990. *Identity and Discrimination*, Oxford: Basil Blackwell.
--1994. *Vagueness*, London: Routledge.
--1999a. ‘Vagueness and Ignorance’, in *Vagueness: A Reader*, Rosanna

Keefe and Peter Smith (eds.), Cambridge: The M.I.T. Press.
--1999b. ‘Vagueness, Truth and Logic’. In *Vagueness: A Reader*. Rosanna Keefe and Peter Smith (eds.). Cambridge: The M.I.T. Press.

Williamson, Timothy and Sainsbury. R.M. 1998, ‘Sorites’, in *A Companion to the Philosophy of Language A Companion to the Philosophy of Language*, Bob Hale and Crispin Wright (eds.), Oxford: Basil Blackwell.

Wright, Crispin. 1999. ‘Further Reflections on the Sorites Paradox’, in *Vagueness: A Reader*, Rosanna Keefe and Peter Smith (eds.), Cambridge: The M.I.T. Press.

Sainsbury, R.M. 1999. ‘Concepts without Boundaries’, in *Vagueness: A Reader*, Rosanna Keefe and Peter Smith (eds.), Cambridge: The M.I.T. Press.

The Ravens

Here is an apparent paradox. (a) “All ravens are black” is logically equivalent with (b) “all non-black things are non-ravens”. One can confirm (b) by citing red carpet fibers and green automobiles. Since (a) and (b) are logically equivalent, it follows that those same carpet fibers and automobiles

equally confirm (a). But this seems very strange. People have dealt with “the paradox of the ravens” either by countenancing non-classical logics or by saying that red carpet fibers *do* confirm (a) as much as they confirm (b) (the appearance to the contrary is explained away as some kind of epistemic epiphenomenon). Neither position has to be taken. We simply have to look beneath the surface of the sentences involved. (a) expresses a nomic, not an accidental, generalization; it expresses a nomic connection between the property of being a raven and that of being black. (b), on the other hand, does not express a nomic connection. There is a causal or constitutive relation between the property of being a raven and that of being black, but there is no such relation between the property of being non-black and that of being a non-raven. Although there is *an* interpretation of (a) and (b) on which they are logically equivalent, that is not the interpretation that is usually intended by utterances of (a). The only interpretation on which (a) and (b) are logically equivalent is the artificial reading whereby (a) expresses a purely extensional proposition, in the same class as the infamous “every coin in my pocket is a quarter”. Once we peel away surface-structure of (a), and delineate its nomic content, the paradox disappears. Attempts to show that our analysis is question-begging or otherwise innocuous are shown, first, to embody the absurd idea that *x causes y* is logically equivalent with *y causes x* and, second, to presupposes

For any two logically equivalent hypotheses, any given piece of evidence will confirm the one hypothesis exactly as much as it confirms the other. In any case, it is hard to see how this could be false. It is hard to see how logically equivalent propositions could receive different degrees of support from a given piece of evidence.

But this seemingly unexceptionable thesis seems to have a paradoxical consequence. Consider:

(R1) All ravens are black.
(R2) All non-black things are non-ravens.

R1 and R2 are logically equivalent. I can confirm R2 by looking at red carpet fibers, green leaves, and white kitchen tiles. Therefore I can confirm that all ravens are black by looking at green bars of soap and white kitchen tiles. But intuitively this seems quite wrong.[31]

Many have argued that, strictly speaking, one *can* confirm R1 by looking at red carpet fibers and green bars of soap.[32] Others have denied that this can be done, and have tried to solve the paradox by appealing to non-standard logics.[33]

Both of these "solutions" are desperate, and neither need be adopted. There is a much simpler solution.

Let us refer to *any* paradox of the type just described as a "raven-type" paradox. Of course, most raven-type paradoxes have nothing to do with ravens. It will help if we initially focus on some of these other paradoxes: the fallacious reasoning that is involved, in a very obscure way, in the original raven-paradox is involved in a very transparent, and easily identified, way in some of these others. So before we give our solution, let us consider two other raven-type paradoxes:

Consider:

(H1) For any x, if x is a human who drank massive quantities of arsenic, then

x is dead.
(H2) For any x, if x is not dead, then x is not a human who drank massive quantities of arsenic.

Suppose that H1 and H2 are logically equivalent. In that case, any evidence confirms the one hypothesis exactly as much as it confirms the other. But this has a paradoxical consequence. We can confirm H2 by finding a person who is *not* dead who has *not* drunk arsenic. Actually, *anything* that is not dead, so long as it is not a person that has consumed arsenic, constitutes support for H2. But it is very hard to believe that I am confirming H1 by citing cases of living people who have not drunk arsenic, or cases of non-dead non-humans that may or may not have drunk arsenic.

Before we give our solution, here is one last example of a raven-type paradox:

(C1) All copper conducts electricity.
(C2) All non-conductors (of electricity) are non-copper.

C1 and C2 are logically equivalent. C2 is confirmed by the fact that the number five is a non-conductor that is non-copper. Thus C1 is confirmed by that same fact. But that is surely wrong.

If I am not mistaken, these paradoxes are easily dissolved. Let us focus on H1 and H2. When somebody puts forth H1, what is he really saying? He is saying something roughly along the following lines: first, all humans that have drunk arsenic are in fact dead; second, that fact is not mere happenstance, but has its basis in some kind of *nomic* or principled connection holding among the properties in question. It is not an accident that everyone who has consumed megadoses of arsenic is no longer living.

Obviously there is some kind of principled connection between the property *human who has consumed megadoses of arsenic*, on the one hand and, on the other hand, the property *no longer living.* There is obviously a *causal* relation between x's having the first property and x's having the second. So when somebody puts forth (H1), what he is really saying is something at least vaguely like this:

($H1_C$) Given any x, if x is a human being, then x's drinking arsenic *causes*[34] death.

H2 is *not* logically equivalent to ($H1_C$). This is a matter of elementary logic. ($H1_C$) says that a person's drinking arsenic causes his death. H2 doesn't say anything about causal relations. Further, there is no way obvious way to reconfigure H2 so as to make it equivalent with ($H1_C$). Consider:

(H2*) For any x, x's being not dead causes x to be something other than a human who drank massive quantifies of arsenic.

H2* is false (in fact, it is arguably absurd); and it is *not* logically equivalent with ($H1_C$). The two aren't even materially equivalent.

Let us generalize the points just made. Speaking very roughly, "all" statements usually affirm relations between classes or between properties. "All politicians are liars" makes a statement about the relationship between the class of politicians and the class of liars. "All copper conducts electricity" makes a statement about the relationship between the class of copper objects and the class of objects that conduct electricity. (There is a nuance here. Depending on how the "all" is read in such statements, they can either affirm a relation between classes *or* they can affirm a relation between *properties.*

This is a point we will often revisit.)

Oftentimes, when people make "all" statements, they are not *just* saying that every object falling into one class falls into some other class. When you say "All copper conducts electricity", you are not *just* saying that, as it happens, all the copper things that exist also conduct electricity. You are making a deeper statement. You are saying that there is some kind of *nomic connection* between the property of being coppery and the property of being a conductor.[35] When you say "any person who just drank a gallon of arsenic is now dead", you are not saying that, as it happens, anything having the first property also has the second. You are making a deeper statement. You are saying that there is some kind of *nomic* connection between, on the one hand, the property of being a person who just drank a gallon of arsenic and, on the other hand, the property of being dead.

"All copper conducts electricity" does not affirm an *accidental* generalization. It affirms some kind of *nomic connection* or, in any case, some kind of principled connection between two properties.

As it would usually be meant, the same is true of H1: it affirms some kind of *nomic* or, in any case, principled connection between two properties. The connection is: having the first property *causes* a thing to have the second property.

I do not wish to be misunderstood. In *some* cases, "all" statement do affirm "accidental" generalizations. The stock example is:

(CP) "For all x, if x is a coin in my pocket, then x is a quarter."

Here all that is being said is that everything falling into one class *happens* to fall into another; everything having one property (that of being a coin in

my pocket) *happens* to have another (that of being a quarter).[36]

Conceivably, one *could* use H1 to express a purely accidental generalization. But that is probably not how it would usually be meant, and it is not how it would usually be taken. Suppose a three year old child is about to drink arsenic. You tell him: “everyone who drinks arsenic dies”. Here what you are saying, and what you will probably be *taken* to say, is that there is some kind of connection – not a purely accidental or extensional one -- between the property of drinking arsenic and that of subsequently dying.

It is a simple fact that, in at least some cases, “all” statements are intended to express nomic – and not merely accidental or extensional – relations between classes. Whether this is a matter of semantics as opposed to pragmatics – of literal meaning as opposed to speaker’s meaning, of explicature as opposed to implicature – is a question we needn’t consider: we can leave that problem to the philosophers of language. We are concerned with what is *communicated* and *meant* (in the psychological sense) by “all” statements, not (necessarily) with what they literally mean. For our purposes, it is enough to know that, if you utter (H1), you are probably making some kind of *nomic* statement; you are probably saying, or at least insinuating, that there is some kind of principled connection holding among the properties involved. What you are saying is at least vaguely along the lines of:

($H1_C$) Given any x, if x is a human being, then x’s drinking arsenic will *cause* death.

Clearly ($H1_C$) is not in logically equivalent with:

(H2) For any x, if x is not dead, then x is not a human who drank massive quantities of arsenic.

These are entirely different propositions; they are *certainly* not *logically* equivalent. So there is *no* reason why a given piece evidence might give more support to the one than to the other.

Raven-type paradoxes always involve the supposed equivalence of two "all" statements: two statements having the form "all phi's are psi's" or "for any x, if x has phi, then x has psi". The key to dissolving this paradox is to realize that at least one of those statements expresses some kind of *nomic* connection --- or, in any case, some kind of explanatorily significant or otherwise principled connection -- between the properties in question. Once this is realized, it becomes impossible to see the second one as logically equivalent with the first.

Here is another example.

(B1) All people who have recently been repeatedly beaten with hard objects are injured.
(B2) All non-injured things are not people who have recently been repeatedly beaten with hard objects.

(B1) and (B2) are logically equivalent (so we are told). I can confirm B2 by citing cases of people who have *not* been beaten or, indeed, by citing cases of *anything* – whether it has been beaten or not – that is not injured and is also not a person. So I can confirm (B2) by pointing to any non-injured tree or to any non-injured person who has not been beaten; therefore I can confirm (B1) by pointing to some non-injured tree or non-beaten and non-injured person.

But surely I *cannot* confirm (B1) by citing non-injured trees or by citing non-injured *non*-beaten people. Yet B1 and B2 are logically equivalent; so

anything that confirms the one should equally confirm the other. A paradox!

But there is no paradox; we just have to make sure that we read B1 the right way. Obviously (B1) is meant to express to express some kind of *nomic* or *causal* statement. (B1) says two things: first, as a matter of fact, anything x that is a human that has been repeatedly beaten with a hard object *is* injured; second, that fact is not mere happenstance, but has its basis in some kind of principled or nomic relation holding among the properties in question. So (B1), as it is usually meant, bears something at least vaguely like the following proposition:

($B1_C$) x's being a person who has recently been repeatedly beaten with hard objects is *causative* of x's being injured.

($B1_C$) is not equivalent with B2. Indeed, B2 is not, so far as I can tell, even *elliptical* for anything that is even remotely close to being equivalent with ($B1_C$). So there is no way to generate the paradox just cited.

Notice that to solve the paradox, we *don't* have change our system of logic – we *don't* have to countenance a deviant system of logic.[37] And we *don't* have to embrace the unpalatable idea that one can confirm B1 by noting that my desk is an uninjured, non-beaten, non-human or by noting that I am not injured and am a human who has *not* been beaten. All we must do is look beneath the surface-structure of B1. Nothing in the surface structure refers to causes or to nomic or otherwise significant connections among properties. But, notwithstanding that fact, when we look at what is conveyed by B1, we find that what is being said[38] is not purely extensional, and does have some kind of nomic or causal or, in any case, explanatorily significant content.

But we are not home free. Consider:

(P1) All pinewood floats (for any x, if x is a piece of pinewood, x floats).
(P2) All non-floaters are non-pinewood (for any x, if x doesn't float, then x isn't pinewood).

Supposedly, P1 and P2 are logically equivalent. I can confirm P2 by noting that lead weights don't float. But that fact surely doesn't confirm P1. So, evidently, a datum can give different degrees of support to logically equivalent propositions. A paradox.

We dissolved similar paradoxes by noting that the first "all" statement expressed a *causal* connection between two properties. I don't think that this exact consideration will dissolve this particular paradox. I don't think that x is a piece of pinewood at time t, and that fact *causes* x to be a floater at time t*. If something is a piece of pinewood at time t, then it is a floater at time t. It isn't as though x has the one property first, and the other property later. So we don't exactly seem to be dealing with a *causal* connection between the two properties. Thus, regrettably, we cannot apply our earlier solution to this paradox.

But the basic principle beneath our earlier solution still applies. P1 doesn't just say that, as a matter of happenstance, everything that is pinewood floats. It seems to affirm some kind of *nomic* connection between the two properties: not a causal connection exactly, but a nomic connection no less. That nomic connection is, at least in its general outlines, given by some statement like:

(P3) For any x, if x is a piece of pinewood, then *in virtue of that fact* x floats.

Another possible way of capturing that nomic connection is with the statement:

(P4) For any x, if x is a piece of pinewood, a *consequence* of that fact is that x is a floater.[39]

It seems to me that each of (P3) and (P4) is *not* logically equivalent with:

(P2) All non-floaters are non-pinewood (for any x, if x doesn't float, then x isn't pinewood).

P2 doesn't say that possession of one property is a *consequence* of possession of another; it doesn't say that, for some properties P1 and P2, a thing has the one *in virtue of* its having the other. The concepts of "consequence" and "in virtue of" seem to be absent from P2. P2 registers a purely extensional or accidental connection; it doesn't make an "in virtue of" statement. Thus P2 is not logically equivalent with P3 or P4. Thus P2 is not logically equivalent with P1. For P1 is equivalent (at least roughly) with P3 or P3 or, at any rate, with *some* "in virtue of" statement.

Thus, if we read P1 the right way, it turns out *not* to be logically equivalent with P2; and the paradox in question is dissolved. The key is to see that P1 alleges some kind of principled connection between two properties: a nomic connection (albeit a non-causal one) or, at any rate, some kind of explanatorily significant or otherwise principled connection.

But our work is not done yet. For there is a compelling objection to what we just said:

Maybe P2 *is* equivalent to some "in virtue of" statement. Maybe P2 *does* allege that things have one property in virtue of having some other property – that x's having one property is a *consequence* of x's having some other property. You are much too quick to suppose that P2 is a mere accidental generalization. It seems quite plausible to suppose that it *does* affirm some kind of nomic or, at any rate, principled connection between two properties. Perhaps once it is made clear what that connection is, the paradox in question recrudesces.

Actually, this objection has a certain validity. I don't think it carries the day. But it does force us to clarify our analysis.

Let x be an arbitrary piece of pinewood. For something to be pinewood is, presumably, for it to have certain properties: in particular, certain cellular and molecular properties. Let $p_1 \ldots p_n$ be these properties. So x's being pinewood is identical with, or supervenes on, x's having $p_1 \ldots p_n$ (or, at any rate, with its having some sub-set of those properties).

For some i, x's having p_i, coupled with x's being placed in water, is *causative* of x's floating (i.e. is preventative of x's sinking). So among the properties *constitutive* of pinewood is some property p_i such that a thing's having p_i is causative of that thing's behaving in certain ways in certain contexts. If x has p_i, then a *causal* consequence of that fact is that x floats in water.

Let us sum up. Suppose x is pinewood. In that case, x floats, and it floats *in virtue* of its being pinewood. What does the "in virtue of" mean in this particular context? For a thing to be pinewood is for it to have (some of) $p_1 \ldots$

p_n. Possession of some of these properties is *causative* of a thing's behaving in certain ways in certain contexts; in particular, it is *causative* of a thing's floating when placed in water.

So P1 turns out to be equivalent to something more or less like:

($P1_C$) Let $p_1 \ldots p_n$ be the properties constitutive of pinewooddom (i.e. of being a thing x such that x is pinewood). For any x, if x is a piece of pinewood, then for some i, x's having p_i is causative of x's floating (i.e. is causative of x's not sinking) when placed in water.

Given this, let us revisit:

(P2) All non-floaters are non-pinewood (for any x, if x doesn't float, then x isn't pinewood).

There is *no* reading of (P2) on which it is even remotely close to being equivalent to ($P1_C$). To see this, let us try to generate such a reading.

($P2_1$) Let $p_1 \ldots p_n$ be the properties constitutive of pinewooddom. For any x, if x is a non-floater, then for some i, x's having p_i is causative of x's not being pinewood.

($P2_1$) is not logically equivalent to ($P1_C$). They make entirely different statements. ($P1_C$) says that having p_i causes a thing to float when placed in water. ($P2_1$) says that having p_i causes a thing to be (or perhaps become) pinewood. Those are entirely different claims. For any x, "x causes a thing to float" is *not* logically equivalent with "x causes a thing to be (or become) pinewood".

Let us take stock. P1 seems to be equivalent with ($P1_C$) or ,at any rate, with something similar to ($P1_C$). Given this, if P2 is taken as a purely extensional statement – a statement that does not affirm any causal relation -- then P2 is definitely *not* logically equivalent with P1. So if we are to generate a raven-type paradox, then we must read P2 as affirming some kind of causal or nomic connection. The obvious reading is ($P2_1$). But ($P2_1$) is not even close to being logically equivalent with ($P1_C$) or, therefore, with P1.

Of course, there are other readings of P2 on which it expresses a nomic or causal relation. Consider:

($P2_2$). Let $p^*_1 \ldots p^*_n$ be the properties constitutive of non-floaterdom (of being a thing x such that x doesn't float). For any x, if x is a non-floater, then for some i, x's having p_i is causative of x's not being pinewood.

It goes without saying that ($P2_2$) is not even close to being logically equivalent to ($P1_C$) or, therefore, to P1. ($P1_C$) and ($P2_2$) both make causal statements. But they make entirely different causal statements.

Perhaps there are other legitimate readings of P2 on which it makes a causal statement. But, I guarantee, on any one of them, P2 will *not* turn out to be logically equivalent with ($P1_C$) or, therefore, with P1. It would be a tedious (and, I think, unnecessary) exercise to go through each of these possible readings. I invite the reader to verify my conjecture for himself.

Let us sum up. P1 doesn't affirm an accidental generalization; it affirms some kind of "in virtue of", or "is a consequence of", relation between properties. So P1 means something like:

(P3) For any x, if x is a piece of pinewood, then *in virtue of that fact* x floats.

or perhaps something like:

(P4) For any x, if x is a piece of pinewood, a *consequence* of that fact is that x is a floater.[40]

When we delineate the meanings of the terms “in virtue of”, and “consequence”, as they occur in these statements, it turns out to be very hard to identify any legitimate reading of P2 on which it is even remotely close to being equivalent with P1.

When we read it the right way, P1 is *not* logically equivalent with P2. We don’t have to invent deviant logics to break the equivalence. We just have to note that, in P1, underneath the innocuous seeming “all”, is a non-innocuous nomic or causal connection. P1 affirms the existence of some kind of *connection* between two properties. When we make it clear what that connection is, it is very hard (impossible, I think) to generate any reading of P2 whereby it is logically equivalent with P1.

We still have more work to do. The words “in virtue of” are used in many different contexts; and they denote many different relations. In the statement “if a thing is a piece of pinewood, then *in virtue of that fact*, it floats when placed in water”, the words “in virtue of” denote one relation. We tried to delineate that relation above. The words “in virtue of” do not *always* denote that relation. Consider the statement:

(T) If x is a triangle, then *in virtue of that fact* x has three sides.

Presumably, the words “in virtue of” in (T) don’t denote quite the relation that they denote in (P3). In P3 they denote a causal relation; in T they denote a logical relation.

Similar remarks apply to the relation between P3 and

(T*) If x is a triangle, a *consequence* of that fact is that x has three sides.

In T*, the term "consequence" denotes a purely logical or mathematical relation; it does not denote a causal relation.

There is, at least arguably, a non-causal *and* non-logical reading of "in virtue of". Consider: [41]

(W) if x is a specimen of H_2O, then *in virtue of that fact* x is a specimen of water.

In (W), the "in virtue of" seems to affirm some kind of *identity*: the property of being water is *identical with* the property of being H_2O. The "in virtue of" thus seems not denote a causal relation. Presumably, it doesn't denote a purely logical relation either.

Exactly similar remarks apply to the relation between P3 and

(W*) if x is a specimen of H_2O, then a consequence of that fact is that x is a specimen of water.

In W*, the word "consequence" does not denote a causal or a logical relations; it denotes what might be called a purely metaphysical relation.

I grant that the expressions "in virtue of" and "consequence" have many meanings. I also grant that they do not always, or even usually, have the meanings they have in P3 and P4. But that doesn't affect the universality of our method of dealing with raven-style paradoxes. It means only that we must apply that method with some discretion. In every case, "in virtue of" denotes

some kind of explanatorily significant relation. If we keep that fact in mind, we prevent the generation of raven-style paradoxes.

Consider:

(W1) For all x, if x is a specimen of water, then x is a specimen of H2O.
(W2) For all x, if x is (a specimen of) non-water, then x is (a specimen of) non-H2O.

Supposedly, W1 and W2 are logically equivalent. W2 is confirmed by the fact that the number five is not H2O, that my desk is not H2O, and so forth. But the scientists who first put forth W1 rightly would have regarded those facts as irrelevant to their hypothesis. So a piece of evidence can have one degree of bearing on a hypothesis and a different degree of bearing on a logically equivalent hypothesis. A paradox.

But there is no paradox. We need only be clear on what W1 is really saying. When chemists first put forth W1, they were not affirming an accidental generalization; they were not saying that, as a matter of sheer happenstance, all specimens of water are specimens of H2O. It seems to me (here I will run into some resistance) that W1 says something along the lines of:

($W1_{NK}$) The natural kind *water* is identical with the natural kind *H2O*.

If this is the right reading of W1, then W1 is not even remotely close to being equivalent to W2. W2 doesn't say anything about natural kinds. Surely W2 doesn't mean:

($W2_{NK}$) The natural kind *non-H2O* is identical with the natural kind *non-*

water.

For “non-water” and “non-H2O” are not natural kinds at all. This must be understood aright; properties are to be distinguished from natural kinds – not all properties correspond to natural kinds (though some do). Consider the property of being a thing x such that (either x is human or x is a square circle). I have that property, and so do you. But that property does not correspond to a *natural kind.* The property of being human is (or generates) a natural kind. But not the property of being *either human or a square circle.*

Some things have the property of being non-water. The number five has that property; so does my desk. But a thing does not belong to a natural kind in virtue of having that property.

By contrast, a thing *does* belong to a natural kind in virtue of being water (H2O). The property of being water generates a natural kind, and so (therefore) does the property of being H2O.

Permit me to make some very approximate statements.

When we say that a property P is associated with a natural kind we are saying (inter alia) that P is explanatorily significant: P is the kind of property used in nomic generalizations and scientific explanations. The property of being water is such a property; it is explanatorily, and predictively, significant. The property of being *non*-water does not have that significance, at least not to a high enough degree to warrant its being a natural kind.

Suppose I tell you: “x is water.” Given a reasonable backlog of scientific knowledge, you could on that basis alone infer an enormous amount about the properties of x. The same is true if I say “x is H2O”. But if I say “x is non-water”, or “x is non-H2O”, you can infer absolutely nothing non-trivial. For all you know, x could be the number five or Alpha Centauri or Bertrand Russell’s favorite line from *Hamlet.* Statements of the form “x is water” and

“x is H2O” integrate with empirical statements in a very different way from statements of the form “x is non-water” and “x is non-H2O”. It isn’t easy to pinpoint *what* that difference is exactly. But, unmistakably, there is *a* difference; and that difference is appropriately described by saying that *water* and *H2O* are natural kinds, while *non-water* and *non-H2O* are not.

I am not denying that statements like “x is non-water” and “x is non-H2O” are significant. Obviously they are replete with significance. And obviously they have meanings that are crucial in many scientific contexts. But “non-water” and “non-H2O” do not denote natural kinds. The statement (or statement-function) “x is either lead or cork” may be of paramount significance in many scientific contexts. But the property “is lead or cork” does not pick out a natural kind.

So

($W1_{NK}$) The natural kind *water* is identical with the natural kind *H2O*.

and

($W2_{NK}$) The natural kind *non-H2O* is identical with the natural kind *non-water*.

are not logically equivalent. They aren’t even materially equivalent. ($W1_{NK}$) is true, whereas ($W2_{NK}$) is false. ($W2_{NK}$) is false because it says, or at least conveys by implicature, that *non-water* and *non-H2O* are natural kinds. But they are not.

So

(W1) For all x, if x is a specimen of water, then x is a specimen of H2O

(W2) For all x, if x is (a specimen of) non-water, then x is (a specimen of) non-H2O.

are *not* logically equivalent. W1 is really a way of saying:

($W1_{NK}$) The natural kind *water* is identical with the natural kind *H2O*.

So W1 says something true about natural kinds. W2 does not say anything about natural kinds; in any case, it doesn't say anything true about them. So, read the right way, W1 and W2 are *not* logically, or even materially, equivalent. Thus the raven-type paradox associated with them cannot even be generated.

Let us now consider the original raven-type paradox:

(R1) All ravens are black.
(R2) All non-ravens are non-black.

This *particular* raven-type paradox is actually one of the harder ones to deal with. (This, I think, has retarded progress on this particular problem.)

First of all, R1 does not affirm an accidental generalization; it is not in the same category as "all coins in my pocket are quarters".[42] So R1 says that, as a matter of fact, everything that is a raven is black *and this is because of some connection between the property of being a raven and the property of being black.* The question is: what is that connection?

For something to be a raven is for it to have certain properties: in particular, certain genotypic and phenotypic properties. Let $r_1 \ldots r_n$ be these properties. Let x be some raven. x is a raven in virtue of the fact that it has (some of) $r_1 \ldots r_n$. x is black; and this fact is connected with its being a raven.

What is that connection? One possible answer is: for some i, r_i is the property of being black. If that is right, then the property of being black is *constitutive* of ravenhood. My own view is that this is not the case; I don't think it is quite correct to say that blackness is constitutive of ravenhood. But let us explore this line of thought for a moment. If the position just described is correct, then R1 really means this:

($R1_{CN}$) blackness is *constitutive* of ravenhood.

First, I should clarify my use of terms. When I say that blackness is "constitutive" of ravenhood, I obviously don't mean that being black *suffices* for being a raven. The property of containing oxygen is constitutive of the property of being water. But for something to be water, it is obviously not enough that it have oxygen in it. *Part* of what it is to be water is to contain oxygen. That is the sense in which containing oxygen is "constitutive" of being water. Henceforth, when I talk about one thing's being "constitutive" of another, I have in mind the meaning just given. Let us now turn to matters of substance.

If ($R1_{CN}$) is the right reading of R1, then R1 and R2 are definitely *not* logically equivalent. R2 surely doesn't say:

(R2) being non-black is *constitutive* of being a non-raven.

R2 strikes me as false. A piece of coal is as much a *non*-raven as a thing could be. Nothing could *more* have the property of being a non-raven than a piece of coal. But coal is quite black. So being *non*-black would appear not to be constitutive of being non-raven.

I should deal with a possible objection:

When you say that property x is *constitutive* of property y, you are really just saying that an instance of x is nomically necessary for an instance of y. Blackness is nomically necessary for ravenhood. By the same token, non-ravenhood is nomically necessary for non-blackness. So non-ravenhood *is* constitutive of non-blackness.

The term "nomically necessary" covers different relations: some causal, some non-causal. Not all nomic connections are causal connections. There is a nomic connection between being copper and having a certain atomic structure. But there isn't a causal connection. It isn't as though a thing is copper first and *then*, as a causal result, it acquires a certain atomic structure. But sometimes the term "nomically necessary" *does* denote a causal relation. When you say that one property is "constitutive" of another, you are not alleging a *causal* relation. The property of being composed of (inter alia) oxygen is constitutive of the property of being water (H2O). But the relation here is not causal.

The term "nomic relation" covers causal and non-causal relations. The relation denoted by "is constitutive of" is narrower. So the statement "property x is constitutive of property y" is *not* always interchangeable with "there is a nomic connection between property x and property y". To say that blackness is "constitutive of" ravenhood is not *just* to say that there is *some* kind of nomic connection between a thing's being black and its being a raven; it is to make a more specific claim; it is to say that, for some set of properties $r_1 \ldots r_n$, for a thing to be a raven just *is* for it to have (some of) $r_1 \ldots r_n$ and, further, that for some i, r_i is the property of being black.

Before we can close our argument, we must deal with a nuance. As we

noted, some properties generate natural kinds, while others do not. Let us refer to the former as NK-properties. So *rabbit, water, raven, oxygen* are NK-properties, while *thing x such that x is either a square circle or is Russell's favorite vegetable* is not an NK-property.

Obviously some NK-properties are built out of others. The property of being a raven or a rabbit is not a simple one; being a rabbit consists in having (some of) a long list of NK-properties: in particular, genotypic and phenotypic properties. When you say that possession of some property p is "constitutive of" ravenhood, what you are saying is not merely that having that property is logically necessary for being a raven. For x to be a raven, it is logically necessary that x not be a square circle, that x not be an even prime greater than two, that x not be a book that Smith both has and has not read. But surely the property of being a thing x such that x is not a square circle is not appropriately described as "constitutive of" ravenhood. Even though it is true, for all x, that *x is a raven* implies *x is not a square circle*, nonetheless it seems false to say that ravenhood is constituted by (inter alia) the property of not being a square circle. And the reason it seems false, I would suggest, is that the property of not being a square circle is not an NK-property. When we talk about the properties that are "constitutive" of ravenhood, we are talking about NK-properties. Of course, the same is true *mutatis mutandis* of the properties that are constitutive of NK-properties like *copper, rabbit, water,* and *pinewood*.

Now we can close the argument.

($R1_{CN}$) blackness is *constitutive* of ravenhood

is *not* equivalent with :

($R2_{CN}$) non-blackness is *constitutive* of non-ravenhood.

Let $r_1 \ldots r_n$ be the properties constitutive of ravenhood. ($R1_{CN}$) says that, for some i, r_i is the property of blackness.

Let $b_1 \ldots . b_n$ be the properties (if any there be) constitutive of non-blackness. R2 says: for some i, b_i is the property of non-ravenhood.

$R2_{CN}$ is *not* logically equivalent with ($R1_{CN}$). It is absurd to say that *what it is* for a thing to be black is for it to be a non-raven. Non-ravenhood is *not* constitutive of non-blackness. Non-ravenhood is not an NK-property (neither is non-blackness). Thus non-blackness is not constitutive, in any relevant sense of the word, of any property.

It is not absurd (though I think it is false) to say that what it is for a thing to be a raven is (inter alia) to be black. It *is* absurd to say that non-ravenhood is constitutive of non-blackness. The relation denoted by "is constitutive of", at least as it is occurring in this context, is one that holds among natural kinds. Non-blackness and non-ravenhood are not natural kinds. So in saying that non-ravenhood is "constitutive of" non-blackness, one is saying something either absurd or ill-formed. By contrast, when one says that blackness is constitutive of ravenhood, one is not saying something ill-formed or absurd (though arguably – see below -- one is saying something false). To sum up, if R1 says that blackness is *constitutive* of ravenhood, then R1 and R2 are *not* logically equivalent; and the paradox of the ravens cannot even be generated.

But we are still not home free. We've seen that *if* ($R1_{CN}$) is the right reading of R1, then R1 and R2 are not logically equivalent, blocking the paradox. But, for reasons I will go into shortly, I rather doubt that ($R1_{CN}$) is the right reading of R1. I don't think it is quite correct to say that blackness is actually *constitutive* of ravenhood. Such a statement is close to true, but not

quite true.

R1 *does* affirm some kind of nomic or, in any case, explanatorily significant connection between the properties of ravenhood and blackness: and this fact *is* instrumental to block the raven-paradox. But the just mentioned connection is not exactly one of constitution. My guess – I am not a biologist, so it is only a guess – is that the connection given by a statement at least vaguely like this:

(*) Let $r_1 \ldots r_n$ be the properties constitutive of ravenhood. Let x be some raven. x is a raven in virtue of the fact that it has (some of) $r_1 \ldots r_n$. x is black; and this fact is connected with its being a raven. What is that connection? The connection is this: for some i, a thing's having r_i *causes* it to be black. One of the properties constitutive of ravenhood is such that there is a causal connection between a thing's having that property and its being black.

Let me explain the motivation for (*). For the sake of argument, let us suppose that, as a matter of fact, all existing ravens are black. It is pretty clear that R1 is not an accidental generalization. I don't know if it is nomically possible for a raven to be non-black. But my guess is: it is possible. We could imagine a raven lacking the pigmentation needed to have black feathers. That doesn't mean that there is no causal, or otherwise nomic, connection between being a raven and being black. Causal connections are sometimes conditional. (Wet roads don't *always* lead to car crashes, but it is often correct to say "a cause of the crash was the wetness of the road".) The causal connection between a thing's being a raven and its being black is of the conditional kind (not the kind that involves partial differential equations).

But if being black were actually *constitutive* of ravenhood – if for a thing to be a raven just *were* (inter alia) for it to be black – then it would not be

nomically, or even metaphysically, possible for a thing to be a non-black raven. So I don't think that $R1_{CN}$ gives the right reading of R1. The right reading of R1 is something at least approximately like this:

($R1_{CC}$) Let $r_1 \ldots r_n$ be the properties constitutive of ravenhood. For some i, a thing's having r_i is causative of a thing's being black. That is why, in point of fact, all ravens are black.

There is *no* legitimate reading of R2 whereby it is logically equivalent to ($R1_{CC}$). Given that ($R1_{CC}$) is, at least approximately, the right reading of R1, what would the corresponding reading for R2 be? Presumably it would be:

($R2_{CC}$) Let $b_1 \ldots b_n$ be the properties constitutive of non-blackness. For some i, a thing's having r_i is causative of a thing's being a non-raven. That is why, in point of fact, all non-black things are non-ravens.

($R1_{CC}$) may well be true. But ($R2_{CC}$) is simply absurd. First of all, for reasons discussed earlier, it probably doesn't even make sense to talk of the properties that are "constitutive", in the relevant sense, of non-blackness. Second of all, even if there are such properties, they surely don't *cause* a thing to be a non-raven. My shoes are non-black. My shoes were not *caused* to be non-black by their being non-ravens; and they weren't caused by to be non-ravens by their being non-black. It isn't as though they were non-ravens first, and then, as a causal consequence, they *became* non-black. ($R2_{CC}$) is obviously ludicrous. By contrast, ($R1_{CC}$) is not ludicrous; indeed, give or take some subtleties relating to the notion of causality, ($R1_{CC}$) is probably true. R1 is, at least approximately, equivalent with ($R1_{CC}$). R1 is *not*

logically equivalent with ($R2_{CC}$); they are probably not even materially equivalent. So R1 and R2, read the right way, are *not* logically (or even materially) equivalent.

Of course, there are readings of R2 other than ($R2_{CC}$). But on none of these reading will the raven-paradox recrudesce. R1, we have seen, holds in virtue of a causal connection between the property of blackness and the property of ravenhood (or, more exactly, one of the properties constitutive of ravenhood). If R2 is to be logically equivalent with R1, then R2 must hold in virtue of a causal connection between the properties of non-blackness and non-ravenhood. But no such causal connection is to be found. A thing's being a non-raven won't *cause* it to be non-black (my shoes are non-ravens and are non-black, but they're being non-ravens didn't cause them to be non-black); a thing's being non-black won't *cause* it to be a non-raven (my shoes are not ravens, but they're being non-black didn't *cause* them to not be ravens); a thing's having the properties *constitutive* of non-ravenhood won't *cause* it to be non-black; and a thing's having the properties *constitutive* of non-blackness won't *cause* it to be a non-raven.

Given that R1 is logically equivalent with $R1_{CC}$, or with something even vaguely like $R1_{CC}$, it becomes impossible to read R2 in a way that makes it logically equivalent with R1. Indeed, it becomes hard to read R2 in a way that makes it *materially* equivalent with R1.

Once we recognize that R1 makes a nomic statement of some kind, and once we identify the right nomic connection, it turns out that R1 and R2 are *not* logically equivalent.

I should deal with a possible reservation about this line of reasoning:

Your solution is basically this. "All ravens are black" (as it is usually understood) is *not* logically equivalent with "All non-black things are non-

ravens" (as it is usually understood). The former alleges a nomic relation, while the latter does not; and once the nomic content in the first is recognized, it becomes hard, or impossible, to see the two as logically equivalent.

But it seems to me that, in this context, it is question-begging to take it for granted that the first does, while the latter does not, have nomic content. What does it mean to say that "all ravens are black" has nomic content? One possible answer is that *it is confirmed by its instances.* "All coins in my pocket are quarters" is not confirmed by its instances: and *that*, plausibly, is exactly what it is for it *not* to have nomic content. So when you say that "all ravens are black" and "all non-black things are non-ravens" are *non*-equivalent, the reason being that the first has nomic content while the latter does not, that is just another way of saying that the two sentences are *confirmationally* non-equivalent. But that is the very thing for which we are seeking an explanation.

The paradox of the ravens *is* that those two sentences (and other pairs like them) are confirmationally non-equivalent. You have "solved" the paradox by saying that the first has nomic content, while the second does not. But that is no solution; it is just a way of saying the first *is* confirmed by its instance, while the second is not. But merely saying that is no solution; indeed, merely saying that is question-begging – you are simply *affirming* the very thing that, under the circumstances, is to be demonstrated.

Here is why I don't accept this. Once again consider (H1) and (H2), and also (B1) and (B2). (H1) and (B1) blatantly make *causal* statements; they are of the form: phi's cause psi's. Of course, *verbally* they are not of that form, but that is scarcely relevant: there can be no doubt as to the form of they are saying. Now nobody will say that *phi's cause psi's* is logically equivalent

with *non-psi's cause non-phi's.* You can take or leave Hume's views on causation; you can take or leave Goodman's and Hume's views on induction. Regardless of your positions on these matters, elementary logic bars you from regarding as logically equivalent *being beaten causes injury* and *being non-injured causes a state of being non-beaten.* In saying this, I am not begging any questions. Anyone who alleged the logical equivalence of *all phi's cause psi's* and *all non-psi's cause non-phi's* would be guilty of a serious logical howler. Once it is conceded that (B1) and (H1) make causal statements, their confirmational *non*-equivalence with (B2) and (H2) flows from basic logic: one is then enjoined by logic *not* see them as equivalent with (B2) and (H2). Not only is it *not* question-begging to see them as *not* logically equivalent; it is logically *de rigueur*.

Also, as we saw, when you say that a statement has *causal* content, that is much more specific than saying that it has some kind of nomic content or other; and it is also much more specific than saying that it receives support from its instances. So, contrary to the objector's point, saying that (B1) and (H1) make *causal* statements – that they are, or have embedded within them, statements of the form *phi's cause psi's* – is simply *not* the same thing as saying that they have nomic content. Of course, saying that (H1) and (B1) have that form certainly does *entail* that they have nomic content; and it also entails that they are confirmed by their instances. But is a far cry from their being merely alternate ways of saying that they are confirmed by their instances. Also, that is exactly what we want. We *want* to explain why (B1) and (H1) do, while (B2) and (H2) do not, receive confirmation from their instances; so we *want* our solution to entail that (B1) and (H1) do, while (B2) and (H2) do not, receive support from their instances. That is they very thing we are trying to explain. When we point out that the first two make causal statements, we immediately establish why the first two are confirmationally

non-equivalent with the second; and we do so by citing some more fundamental fact from which that non-equivalence is derived. The fact that the pairs in question are non-equivalent is obviously a *consequence of* (though by no means identical with) the fact that the first pair consists of causal statements, while the second do not.

There is more to say in response to the objector's point. As we've seen, (P1) probably doesn't say *being pinewood causes a thing to be a floater*, and (R1) probably doesn't say *being a raven causes thing to be black.* But we've seen that, this fact notwithstanding, (P1) and (R1), when delineated, prove to have *causal content*: tucked within them are statements of the form *phi's cause psi's.* So even though the nomic connection between ravenhood and blackness, and between pinewooddom and floaterdom, is not exactly given by a statement of the form *phi's cause psi's*, nonetheless embedded within those nomic relations are causal facts. And those causal facts, as we saw earlier, destroy the putative equivalence between (R1) and (R2), and also between (P1) and (P2).

Let me make this more concrete. We argued that (R1) is really a way of saying ($R1_{CC}$). ($R1_{CC}$) Once that is granted, it becomes hard or impossible to produce a reading of (R1) that is logically even remotely like (R1). The impossibility here is, ultimately, not so different from the impossibility of regarding (B1) and (B2), or (H1) and (H2), as logically equivalent. What we just said about (R1) applies *mutatis mutandis* to (P1). The non-convertibility of causal statements – the fact that *phi's cause psi's* isn't (typically) interchangeable with *psi's cause phi's* (typically they aren't even *materially* interchangeable) – is not open to dispute, and can be presupposed with any questions being begged. And, once that non-interchangeability is conceded – and how can it not? – then we have destroyed *many* (though perhaps not all) raven-type paradoxes.

Nonetheless, we have not yet completely neutralized the objector's point: we have neutralized it only for those raven-type paradoxes that involve *causal* statements. We have argued that many of them do involve such statements. But not *all* of them do: some involve nomic statements that, by all appearances, are *non*-causal. For example, (W1) doesn't say, even implicitly, that anything causes anything else. Now we the confirmational non-equivalence of (W1) and (W2) by saying that the first does, while the second does not, have nomic content. Since (W1) is not on any level a casual statement, it *does* seem possible that, in this case, the objector is right: in saying that (W1) has nomic content while (W2) does not, perhaps we have merely said, albeit in different words, that (W1) is, whereas (W2) is not, confirmed by its instances. And if that is what we have done, then the objector's point stands, at least for that particular raven-type paradox and other raven-type paradoxes of its ilk.

Our solution to this particular problem involved the notion of a natural kind: (W1) says something of the form *natural kind x is identical with natural kind y*; (W2) does *not* say anything about natural kinds. Given this, it is, in my view, clear why the objector's charge of question-begging is not applicable. Embedded within the concept of a natural kind is the concept of *causality*. I am not saying that the concept *natural kind* is a purely causal one. But I would be very surprised if statements like *copper is a natural kind* were causally innocent. When you say that the property of being copper generates a natural kind, whereas the property of being either a square or a non-square does not generate a natural kind, surely *part* of what you are saying concerns the causal relations (if any) mediated by those properties. Surely part of the reason copper is a natural kind is that, if a thing is made of copper, that means it has certain causal powers: if a thing is made of copper, then *ipso facto* it conducts electricity, is malleable to a certain degree, has a certain

melting point, and so on. So, plausibly, the concept of a natural kind is, to a large extent (though perhaps not entirely), a *causal notion.* If you say that something is either a square or a non-square, that is a causally sterile statement. This, presumably, is what underlies the fact that properties like *copper* are explanatorily and predictively rich, and what underlies the fact that properties like *square or non-square* are sterile in those respects. So once we disentangle the logical form of statements like *water is a natural kind,* those statements *would,* I believe, turn out to embody *causal* claims. So statements like ($W1_{NK}$) are, at last in part, causal statements. For reasons we have, in effect, already seen, this would make it impossible to generate a reading of (W2) whereby there is any equivalence, even material equivalence, between it and ($W1_{NK}$) or, therefore, between it and (W1).

To sum up, even though not all nomic statements are purely causal in content, they all turn out to have causal content; and once this is granted, it becomes *de rigueur* (and thus *not* question-begging) to regard as *non*-equivalence the pairs of sentences that generate raven-type paradoxes.

There is another, deeper reservation that I have about the objector's point. To say that (e.g.) *all ravens are black* receives support from its instances is surely not the *same* thing as saying that it has nomic content. In any case, those statements are the same only if one accepts an extremely austere, hyper-empiricized conception of nomic connectedness. The traditional (non-Hume, non-van Fraassen) concept of a nomic connection is something like this. When you say that the lightning causes thunder, or that something about being a raven leads to having black feathers, you are making a statement about relations between things or properties; you are saying that there is some kind of internal (in a broad sense of the word) or essential relation between the property of being a raven and that of having black feathers. The obvious thing to say here is that, if that kind of internal or essential relation is present,

then a *consequence* is that the statement has confirmational properties not had by nomically empty statements. But the idea is that nomic facts *ground* confirmational facts; the latter are parasitic on, and distinct from, the former. Intuitively, w want to say that the moon's causing the tides doesn't *consist* in some proposition's having certain confirmational properties; it *consists*, we feel, in there being some connection in the things themselves.

Of course, no shortage of great thinkers (starting with Berkeley and Hume) have argued for the falsity of what I have been calling the traditional picture. Many have argued that there are no internal connections of the kind described. And, what is related, many have argued that, when you saying things like the *lightning causes thunder*, you are really just making a statement about the confirmational relation between certain bits of data and other bits of data.[43] But, obviously, it cannot be taken for granted that this conception is correct. The orthodoxy these days is to say that it is *not* correct. In any case, the objector's point is conditional on the truth of a, to say the least, exceedingly controversial and counter-intuitive conception of nomic connectedness. If you take a non-Hume-Berkeley line on nomic connectedness – if say you that there is more to nomic connectedness than confirmational relations among data, and that nomic connectedness properly *grounds* such relations, rather than being identical with them – then the objector's point immediately collapses.

But, as we saw, the truth is that, even if you *do* concede the truth of that austere and controversial conception of nomic connectedness, the objector's point still fails. Nobody – whether or not they adopt Hume's or van Fraassen's views on nomic relations and causality – will say that *phi's cause psi's* is convertible with *non-psi's cause non-phi's.* Once this is acknowledged, then the supposed equivalences which create raven-type paradoxes are destroyed, as we saw.

Conclusion

Every raven-type paradox has the following form:

(1) All phi's are psi's.
(2) All non-psi's are non-phi's.
(3) Therefore one can confirm (1) without ever considering so much as a single phi.

But where each raven-type paradox is concerned, the first of those three sentences actually alleges a *nomic* or, at any rate, explanatorily significant connection of some kind. It never alleges a mere happenstance, or purely extensional, connection. *All water is H2O. All copper conducts electricity. All pinewood floats. All ravens are black.* These all generate raven-type paradoxes. None of these are in the same category as the infamous *all coins in my pocket are quarters.* Any "all" statement that generates a raven-type paradox affirms some non-accidental connection. Once we identify the relevant connection, then (for the appropriate values of "phi" and "psi") the *second* of the above three sentences turns out *not* to be logically equivalent with the first. Read the right way, (1) and (2) turn out to *not* be logically equivalent. This results in the falsity of (3). No non-classical logics need to be invoked; and we don't need to swallow the absurd position that I can confirm "all copper conducts electricity" by noting that the number five doesn't conduct electricity.

Ayer, Alfred Jules. 1972. *Probability and Evidence.* New York: Columbia University Press.
Cohen, Jonathan (1989) *The Philosophy of Induction and Probability*, Clarendon Press: Oxford.
Hempel, Carl. G. (1945a) *Studies in Logic and Confirmation.* Mind 54, 1-26.
Hempel, Carl. G. (1945b) *Studies in Logic and Confirmation. II.* Mind 54, 97-121.
Hempel, Carl G. 1966. *Philosophy of Natural Science.* Englewood Cliffs: Prentice-Hall. Falletta, Nicholas. (1983) *The Paradoxicon: a Collection of Contradictory Challenges, Problematical Puzzles, and Impossible Illustrations*.
Gardner, Martin (1973) *Mathematical Puzzles and Diversions*, Penguin Books.
Leavitt, F. (1996). “Resolving Hempel’s Raven Paradox.” Philosophical Inquiry 18, nos 3/4: 116.
Goodman, Nelson. (1954). *Fact, Fiction, and Forecast.* Cambridge: Harvard University Press.
Mackie, J.L. (1977). *The Cement of the Universe.* Oxford: Clarendon Press.
Maher, P. (1999). “Inductive Logic and the Raven Paradox”. *Philosophy of Science* 66: 50-70.
Quine, Willard van Orman. (1969). “Natural Kinds”. In *Ontological Relativity and Other Essays.* New York: Columbia University Press.
Sainsbury R. M. (1997) *Paradoxes*, (2nd Ed.) Cambridge University Press.
Scheffler, Israel (1981) *The Anatomy of Inquiry.* Indianapolis: Hackett.
Schlesinger, G (1974) *Confirmation and Confirmability*, Clarendon Press - Oxford.
Sylvan, R. and R. Nola. (1991). “Confirmation without Paradoxes”, *Advances in Scientific Philosophy: Essays in Honour of Paul Weingartner*, eds. G.

Schurz and G.J.W. Dorn. Amsterdam: Rodopi: 5-44.
Trusted, Jennifer (1979) *The Logic of Scientific Inference*, The MacMillan Press Ltd.

Are Definite Descriptions Referring Terms?

Part I. Introduction

Russell's theory and the purpose of the present inquiry

Let us refer to sentences containing definite descriptions as 'D-sentences'. Russell's 'Theory of Descriptions' concerns the semantics of such sentences. His theory is essentially this: given any sentence of the form

(i) '….the phi….'

that sentence has as its literal meaning

(ii) *There is exactly one phi and…that individual…*

Or more formally,

(ii_F)[44] *There is some x such that given any y, y=x iff y has phi and….x….*

So, for example, the literal meaning of

(iii) ‘The king of France punched Bob’

is

(iv) There is exactly one king of France, and that individual punched Bob

or, more formally,

(iv_F) *There is some x, such that given any y, y=x iff y is a king of France, and x punched Bob.*

In this essay, the distinction between *sentences* and *propositions* will be important. Given this, permit me to make a couple of points that might seem extremely trivial, but that will serve us well later on. First of all, the *meaning* of a sentence is not itself a sentence: the *meaning* of (iii) is not a sentence. So (iv) is not a sentence. But, of course,

(v) ‘There is exactly one king of France, and that individual punched Bob.’

is uncontroversially *a* sentence that has (iv) as its meaning. So Russell’s theory predicts that (iii) and (v) have the same meaning.

Let us say that (v) is the ‘Russellian paraphrase’ of (iii).[45] As I am using the term ‘paraphrase’, *sentences* paraphrase *sentences*. (Propositions do not paraphrase anything.) To generalize a point made a moment ago: it is clear that, if Russell’s theory is right, then any given D-sentence has the same meaning as its Russellian paraphrase.

A problem for Russell's theory

Notice that, if Russell's theory is correct, then in saying 'the king of France is bald' I am *asserting* (inter alia) that there is exactly one king of France. In general, if Russell's theory is correct, then in tokening a sentence of the form '….the phi….' I am *asserting* that exactly one object has phi. Many have the intuition that this is not so: it is often said that, in tokening a sentence of the form '…the phi…' I am not asserting, but merely *presupposing* that there is a phi.[46]

This intuition is not hard to support. If someone said 'there is exactly one king of France and that person is bald', we wouldn't hesitate to say 'you are wrong: there is no king of France.' But, as Strawson pointed out, if someone said 'the King of France is bald', we'd be reluctant to dismiss that person's words as *wrong*; we'd be more inclined to see them as neither true nor false – as an abortive attempt to make a statement. If we assume that 'the king of France' is what it *appears* to be, viz. a referring expression, then it is clear why we would react this way. If 'the king of France' is a referring expression, and it has no referent, then no statement is made by a token of 'the King of France is bald'; *a fortiori* nothing *false* has been said. This, of course, is consistent with our being disposed to regard a tokening of 'the king of France is bald' *not* as straightforwardly false, but rather as abortive – neither true nor false. Of course, if Russell's theory were correct, then a token of 'the King of France is bald' would simply be false, not abortive: and this fact is prima facie inconsistent with our hesitating to regard such a tokening as either true or false.

Russellians try to neutralize this apparent problem for their theory with the following consideration. First of all, contextual factors mediate between

literal and communicated meaning. Consequently, accurate judgments about semantics do not always follow straightaway from speaker's intuitions. For these judgments typically embody a grasp of *communicated* meaning rather than *literal* meaning (semantics). Now given only that 'the kind of France is bald' *seems* to be 'abortive' in a way that its Russellian paraphrase is not, it doesn't follow that its *semantics* isn't what Russell says: some kind of pragmatic hidden variable might be distorting our grasp of the literal meaning of that sentence. And once pragmatics is taken into account, our linguistic intuitions are reconciled with Russell's theory. [47]

The purpose and plan of this paper

In this paper, I wish to show that Strawson is right. Definite descriptions are referring expressions, not quantifiers.

To settle the question whether definite descriptions are quantifiers or means of securing of reference, it will not do to rely on unrefined intuitions; for it is the validity of these very intuitions that the Russellian will call into question. Given this, we need some kind of *test* of synonymy: some way of factoring out the literal meaning of a sentence-token from the non-semantic information in which it may be embedded.

Accordingly, the plan of this essay is as follows. In Part II, I will develop a test of synonymy – a way of separating what is literally meant from what is pragmatically imparted. In Part III, I will show that, for any D-sentence S1, if S2 is the Russellian paraphrase of S1, then the pair (S1, S2) *fails* our synonymy test and that this failure has to do with the fact that S2 *asserts* what S1 merely presupposes.

To say what I will do in Part IV, a bit of background must be supplied. Usually, definite descriptions *seem* to be functioning referringly. The obvious

(and, in my view, the correct) thing to say in such cases is that, at the level of semantics, the definite description *is* a referring term. But there are cases where the obvious analysis is to say that the definite description is *not* a referring term. Consider the sentence "Socrates was the greatest philosopher of antiquity." Here the definite description seems to be functioning as a *predicate*;[JK1] it expresses an attribute.

In part IV, I will show that, at the level of *semantics* or literal meaning, definite descriptions are *always* referring expressions, even though they are often *used* to do things other than refer – even though, for example, they are often used to express a property rather than pick out an object.

In Part IV, I will also show that Russell's theory is not a piece of *conceptual* analysis, but is a *bona fide* semantic analysis. He must be taken to be saying what sentences *mean*, and not just to be providing equivalences of truth-conditions.

One last statement is in order concerning the scope of the present inquiry. Russell's treatment of definite descriptions is really part of a larger project: an attempt to provide a semantics for *quantifiers*: expression like "some man", "three women", "all birds", and so on.[48] I will focus *only* on Russell's treatment of "the" in the singular; I will not deal with his treatment of "some", "all", and so on.

Part II. The synonymy-test

First let us settle some terminological matters. If a sentence S has as its *literal* meaning a proposition P, let us say that S *semantically encodes* P. Oftentimes a sentence will be used to *convey* or *communicate* some proposition other than or additional to its literal meaning. If you and I are robbing a bank, and I say 'the cops are coming', I am trying to *communicate* to you that we should

leave the bank. But the literal meaning of my words is not *we should leave the bank.* If I say to a waiter 'I'd like a vodka tonic', I'm *communicating* an imperative: *bring me a vodka tonic.* But the literal meaning of my words is not an imperative at all.

So we must be careful to distinguish what is semantically encoded in a sentence, or token thereof, from what is *communicated* by such a token: these needn't coincide. Sometimes I will use the word "convey" as a synonymy for "communicate" (as just defined).

The naïve synonymy-test

Clearly what is communicated doesn't always coincide with what is literally meant. But there is obviously *some* connection between what a sentence literally means and what is communicated by tokens of it: what is communicated by a token of a sentence is a function *in part* of that sentence's literal meaning. *Part* of the reason my utterance of 'the cops are coming' communicates the proposition *we should leave the bank* is that it literally means *the cops are coming.* If I had said something with a different literal meaning (e.g. 'the cops are not coming yet' or 'keep your eye on the security guard') I would assuredly *not* have communicated the proposition *we should get out of the bank.*

Given that what a token of a sentence communicates is a function *in part* of its literal meaning, it would seem to follow that, for any two sentences S1 and S2, if S1 and S2 are truly synonymous, then *ceteris paribus* a token of S1 ought to communicate the same proposition as a token of S2. In other words, given that literal meaning is *a* factor in what a sentence-token communicates, it would seem to follow that if all factors *other* than literal meaning are held constant, then an utterance of S1 should communicate the same proposition

as an utterance of S2, provided that S1 and S2 really are synonymous. This is not really a point about semantics specifically; it is just a general methodological principle that is being brought to bear on a semantic question. If x and y are qualitatively identical objects or events, then ceteris paribus – i.e. holding all factors constant (modulo the constitutions of x and y) - x and y should have the same effects. Conversely, if it is *not* the case that tokens of S1 and S2 communicate the same proposition (holding all factors constant, modulo their literal meanings), that would seem to be good evidence that S1 and S2 are not synonymous.

Let us refer to this as our 'naïve synonymy test.' (In a moment I will explain why it is naïve.) So a pair of sentences (S1, S2) *fails* our naïve synonymy test if there is some circumstance C such that, holding all factors constant (modulo the literal meanings of S1 and S2), a token of the one sentence in C will communicate a different proposition from a token of the other sentence.

Supposing that our naïve synonymy test were valid, we could assess the validity of the theory of descriptions in the following way. Let S1 be a D-sentence and let S2 be its Russellian paraphrase. (So S1 might be 'the king of France is bald' and S2 would then be 'for some x, given any y, y=x iff y is a king of France, and x is bald.') If it turns out that the pair (S1, S2) *fails* our synonymy test that proves that they are not synonymous and that Russell's theory fails. But before we apply our test, we must cross a few more bridges.

The problem of tone

There is a problem with our naïve synonymy test. It is commonly held that two expressions, and by extension two sentences, can be synonymous but differ in 'tone'. (Often the term 'connotation' is used instead of 'tone'.)[49]

The term ‘grease monkey’ seems to be synonymous with ‘auto-mechanic’; but these words differ in ‘tone’. By extension, the sentences ‘all grease monkeys are happy’ and ‘all auto mechanics are happy’ are synonymous but differ in tone. (The latter sentence is ‘neutral’; it doesn’t embody any value judgment about auto mechanics. By contrast, the former sentence embodies an attitude of contempt towards them.)

This voids our naïve synonymy test. To see why this is so, once again consider the sentences ‘all auto mechanics are happy’ and ‘all grease monkeys are happy’. Even if *all* contextual factors are held constant, a token of the latter would probably not communicate the same information as a token of the former. For a token of the latter would, while a token of the former would not, communicate the information that the speaker held auto mechanics in contempt. So our naïve synonymy test wrongly says that these two sentences are not synonymous.

One might try to repair our synonymy test by saying: ‘according to that test *all* factors (besides literal meaning) are to be held constant; but you failed to hold their *tones* constant; so you didn’t really apply our synonymy test after all.’ But in response it must be asked: how *could* we hold constant the tones of these sentences? For in order to do that, we’d have to have some criterion by which to distinguish a sentence’s (and, more generally, an expression’s) literal meaning from its tone; and the whole point of our synonymy test is to *isolate* an expression’s literal meaning. So we couldn’t prescind from tone *until* we had isolated the literal meanings of the expressions involved; but it is precisely this that we are trying to do *through* our test.

Solving the tone problem

Even so, there is hope: our synonymy test has not been *completely* voided.

First of all, it seems that not all sentences *have* tone; many are tone-free. A sentence cannot have a tone unless it embodies some kind of emotional, or moral or aesthetic, reaction. But many sentences seem to be completely neutral in this respect, e.g. ‘all auto mechanics are over 6 ft tall’. A being that had no emotions could, without misrepresenting his state of mind, utter such sentences. (By contrast if a completely emotionless being were to say ‘all grease monkeys are over 6 ft tall’[50], he *would* be misrepresenting his state of mind; he would be representing himself as having an emotional attitude of contempt towards auto mechanics when, of course, he could not possibly have such an attitude.) Now tone-free sentences obviously cannot *differ* in tone. It follows that differences in tone cannot account for differences in what is communicated by such sentences.

This gives us a way of reviving our naïve synonymy test. Let S1 and S2 be two sentences. And suppose that there is some context C with the following property. Holding all factors constant, modulo the literal meanings *and* the tones of S1 and S2, the proposition that a token of S1 communicates in C is different from the proposition that a token of S2 communicates in C. In that case, that difference in communicative function is due *either* to semantic meaning *or* to tone. If S1 and S2 are tone-free, then we can rule out tone, and we can conclude that S1 and S2 differ in literal meaning.

This suggests a new strategy. Using our naïve synonymy test, we show that D-sentences differ from their Russellian paraphrases *either* in semantic meaning *or* in tone. But we make sure to apply our test to sentences that do not embody any kind of emotive or valuative attitude – to sentences that have no tone. We can thus conclude that the just mentioned differences are differences in literal meaning.

Another wrinkle

I should respond to a possible objection (made to me in private correspondence by Nathan Salmon):

Might two sentences (or, more correctly, sentence-tokens) that coincide in semantic meaning differ in some respect *other* than tone, e.g. in *emphasis*? Consider the sentences (i) 'all heathers are lovely' and (ii) 'all gorses are lovely'. These have the same semantic meaning and tone (if any two sentences do). But what an utterance of the one communicates may be different from what an utterance of the other communicates, depending on where the emphasis falls in these utterances. If I say (a) '*all* heathers are lovely' that may communicate a slightly different proposition from (b) 'all *gorses* are lovely' or from (c) 'all gorses are *lovely*'. (c) might communicate the proposition 'all gorses are *lovely* but not quite *beautiful*'; (b) might communicate the proposition 'all gorses are lovely, but not all members of some other kind of flower are lovely'; and so on. So it seems that, contrary to what you say, two sentence-tokens can differ in respect of what they communicate, even if those sentence-tokens have the same semantic meaning and tone.

There are a couple of points to make in response to this. First of all, word-emphasis is really a contextual factor. In any case, it is a factor that can be held constant. When we apply our synonymy-test, we will assume that none of the words in the sentences involved are unnaturally emphasized. So to determine whether 'the king of France is bald' is synonymous with 'for some x, given any y, y=x iff y is a king of France, and x is bald', we will assume

that neither of the sentence-tokens contain artificial emphases e.g. 'the king of France is *BALD*' or 'for some x, given any y, y=x iff y is a king of *FRANCE* and x is bald.' For obviously such emphases would introduce a complicating, distorting factor: a factor that alienates us from literal meaning.

It should be pointed out that, in prescinding from such 'artificial' emphases, we are not guilty of some unrealizable idealization: it is quite possible – indeed, it is usually the case – that when a sentence is uttered, it is uttered in an emphasis-neutral way: one says simply 'the man in the corner is a lawyer', not (except to disabuse someone of some gross error) 'the *man* in the corner is a lawyer' or 'the man in the *corner* is a lawyer'. To sum up, emphasis is really a contextual factor that we can prescind from.

The problem of incidentally imparted information

There is one more bridge we must cross before we can apply our synonymy test with a clear philosophical conscience. Consider two sentences that practically[51] *everybody* agrees are synonymous and identical in tone:

(i) Heathers are lovely

(ii) Gorses are lovely.

Even though (i) and (ii) are synonymous and have the same tone, the information that will be imparted by an utterance of the one sentence will not be *quite* the same as the information imparted by an utterance of the other, even if all factors are held constant – even if, for example, those sentences are tokened in qualitatively identical contexts. For example, if I utter (i), my interlocutor is being apprised of the fact that I believe 'heathers are lovely' to

be a true sentence; of the fact that I can produce an ‘h’ sound; of the fact that the word ‘heathers’ is in my vocabulary. *None* of these bits of information would be imparted if I were to utter (ii). So, holding all factors constant (modulo literal meaning and tone), we have a case where what is imparted by a token of one sentence will not be identical with what is tokened by an utterance of a sentence with the same meaning and tone. So we have another challenge to our synonymy test.

This is not an insuperable challenge. If I say ‘grease monkeys make a lot of money’ I am indeed (if only by accident) apprising you of the fact that I can produce a ‘g’ sound. But this bit of information-delivery seems to have an *incidental* character. It is not so much a product, as a *by*-product, of my speech-act. Let us now give a precise statement of the difference between incidentally and non-incidentally imparted information.

I cannot sincerely and *competently* say ‘grease monkeys make a lot of money’ without intending to communicate, first, that I think auto mechanics make a lot of money and, second, that I have some kind of negative attitude towards them. Suppose I sincerely say “grease monkeys make a lot of money”, and I do *not* have an intention to express contempt towards auto-mechanics. In that case, I’m guilty of a *linguistic* failing: a failure to appreciate a dimension of the significance of the expression “grease monkey.” Suppose I sincerely utter that same sentence and do not have an intention of expressing my belief that it is lucrative to be an auto-mechanic. Again, I am guilty of *linguistic* incompetence: of not grasping the literal meaning of my words.

But suppose I say “grease monkeys make a lot of money” and do not have an intention of showing you that I can make a g-sound. *That*, surely, does not show linguistic incompetence. In speaking, people inadvertently – without intending it or even knowing it – show others than they can make g-

sounds. This does not make them guilty of not having a mastery of their language. Of course, if I wish to show someone that I know how to make a ‘g’ sound, I *can* say e.g. ‘grease monkeys make a lot of money’. But I can *competently* use that sentence *without* having any intention at all to apprise my audience of my ability to make that sound.

We can thus define ‘incidental information’ as follows. Suppose I am a competent speaker of English and I token some sentence S; and suppose that *i* is an arbitrary bit of information that is being delivered to my audience as a result of this tokening. If I cannot, *qua* competent speaker of English, token S without intending to communicate *i*, then *i* is non-incidentally imparted information. Otherwise *i* is incidentally imparted information.

It should be noted that incidental information is never *asserted.* If I say ‘snow is white’ I am not *asserting* that I can make an ‘s’ sound. Even if I say ‘snow is white’ with the *intention* of showing you that I can make an ‘s’ sound, I am still not really *asserting* that I can make that sound.

Our test of synonymy (final version)

Suppose, then, that S1 and S2 are synonymous sentences that have the same tone. In that case, provided all factors (modulo tone and meaning) are held constant, what is *non-incidentally* imparted by an utterance of the one must coincide with what is non-incidentally imparted by an utterance of the other.

Our strategy, then, will be this. We will consider D-sentences and their Russellian paraphrases. To simplify matters, we will consider only tone-free sentences. We will see that, even if all factors are held constant, D-sentences and their Russellian paraphrases are not interchangeable: what is communicated by a D-sentence will differ from what is communicated by its Russellian paraphrase *after all non-semantic differences have been allowed*

for. This will leave only semantic differences to blame.

Part III. Why Russell's theory fails the synonymy-test

Consider the sentence:

(a) My neighbor uses my pool frequently.

According to Russell's theory of descriptions, the literal meaning of this sentence is

(b) I have exactly one neighbor and exactly one pool, and that neighbor uses that pool frequently.

It is not hard to think of a context C such that, holding all factors constant, modulo the literal meanings and tones of (a) and (b), the non-incidental information communicated by a token (a) in C would differ from that communicated by a token of (b) in C. So it is not hard to show (a) and (b) are not synonymous.

Suppose that person A works for the census bureau and is interviewing person B:

A: How many neighbors and pools do you have? And what are the salient characteristics of your neighbors?

B: I have exactly one neighbor and one pool, and that neighbor uses that pool frequently.

This dialogue is in no way unusual; neither participant has broken any semantic or pragmatic rules. But consider the following:

A: How many neighbors and pools do you have? And what are the salient characteristics of your neighbors?

B: *My neighbor uses my pool frequently.

This dialogue is obviously peculiar. In a subtle but unmistakable way, B's response is inappropriate; it isn't *quite* responsive to A's question. By contrast, B's statement in the first is dialogue is squarely responsive to A's question. If 'my neighbor uses my pool frequently' were genuinely synonymous with 'I have exactly one neighbor….' – if these two sentences were but notational variants of each other – then B's response in the second dialogue would be every bit as appropriate as his response in the first. To put it another way: if, in saying 'my neighbor uses my pool', I were *asserting* that I had exactly one neighbor and one pool, then B's response in the second dialogue would be unexceptionable. But it isn't.

What exactly has B done wrong? As we just noted, it seems that B has failed to *assert* how many neighbors and pools he has. And because of this, B has violated Grice's 'maxim of quantity': 'you should *assert* neither more nor less than is required for the purpose at hand.' [My emphasis][52] In the second dialogue B seems to be *presupposing* existence where, given the context, he ought to be *asserting* it. Given A's question, it is required of B that he *assert* how many pools and neighbors he has; but B has, by all accounts, failed to do this. If Russell's analysis were correct, then B *would* indeed have *asserted* (inter alia) that he had exactly one neighbor and one

pool; and his response would have been appropriate.

Might this difference be one of *tone*? Both of these sentences seem to be tone-free: neither uses any tone-laden word or (otherwise) embodies any value judgment. So neither has tone; therefore they don't differ in tone.

The difference in question has nothing to do with *incidentally* imparted information. If someone asks me how many neighbors and pools I have, and I sincerely attempt to respond by saying 'my neighbor uses my pool', that shows a lack of linguistic competence: no sincere *competent* English speaker would respond that way. Anyway, incidentally imparted information is (by definition, really) not *asserted*: so the presence of such information couldn't account for assertoric differences between these two sentences.

So neither differences in tone nor differences in incidentally nor imparted information nor differences in contextual factors account for the discrepancy between B's responses in the two dialogues. Given this, it seems we have no choice but to see this discrepancy as indicating a genuinely semantic difference.

Before continuing, I should respond to an objection to what has been said so far:

Once again consider the following dialogue (what we say about it can obviously be extended the other dialogues we have considered):

A: How many neighbors and pools do you have? And what are the salient characteristics of your neighbors?

B: *My neighbor uses my pool frequently.

Admittedly B's response is odd; and it would not have been odd had he said

‘I have exactly one neighbor and exactly one pool and that neighbor uses that pool frequently.’ But this doesn’t prove that Russell’s theory is wrong. For it is easy to prove that the presence of a definite description *does* correspond to an existence claim; and this can be done by very slightly touching up B’s response. Suppose that B had said, not ‘my neighbor uses my pool frequently’, but ‘my *one* neighbor uses my *one* pool frequently’. (The terms ‘unique’ and ‘only’ and ‘one and only’ could probably serve just as well.) In that case, B’s response would have been perfectly appropriate, proving that, just as Russell thought, definite descriptions *do* correspond to existence-claims. [53]

I have a couple of points to make in response. First of all, Russell’s theory is, in effect, that definite descriptions *by themselves* correspond to assertions of existence and uniqueness. So Russell’s theory is, in effect, that ‘my pool’ *by itself* corresponds to an assertion of both existence and uniqueness. So in order to normalize B’s response – to make it non-deviant – it shouldn’t be necessary to *add* some expression like ‘one’ or ‘only’ or ‘unique’.

This point may require some clarification. According to Russell,

(a) my neighbor uses my pool frequently

is synonymous with

(b) I have exactly one neighbor and one pool, and that neighbor uses that pool frequently.

So, if Russell is right, (a) *by itself* asserts that the speaker has exactly one neighbor and pool. So it should be redundant to *add* a term like ‘one’ or

‘unique’ to the definite descriptions occurring therein. In any case, if Russell’s theory were correct, then adding a ‘one’ or a ‘unique’ to the definite descriptions wouldn’t have such a radical effect on the function served by these sentences in the context in question; it certainly wouldn’t change them from *not* being assertions of existence and uniqueness to being assertions of existence and uniqueness: for Russell’s theory just *is* that, even without a ‘one’ being added, (a) is an assertion of both existence and uniqueness.

I would like to make two last remarks on this topic. Even if we leave aside everything that we’ve just said, there is some reason to think that the objector’s remark fails on its own terms. Again consider the following dialogues:

A: Tell me exactly how many Saudi Princes there are.
B: There is exactly one Saudi Prince, and he is a good man.

A: Tell me exactly how many Saudi Princes there are.
B: *The one Saudi Prince is a good man.

The second dialogue strikes my ear as ever so slightly peculiar, whereas the former does not. Both sentences are tone-free. So tone is not to blame. Also the difference obviously doesn’t have to do with contextual factors; even if both dialogues occur in qualitatively identical contexts, B’s response in the second will sound odd, while his response in the first will be normal. Finally, the difference is not due to incidental information: the problem in the second dialogue is obviously endemic to B’s speech-act itself, not to some incidental by-product of it.

So why is B’s response in the second dialogue so peculiar? Surely there is only one thing to say. By using a definite description, even one enriched with

the word ‘one’, B is presupposing existence and uniqueness where, given the context, he ought to be asserting it. Thus we have a counterexample to Russell’s theory and *also* to the objector’s refurbishment of it.

Part IV: Reference versus predication and attribution, semantic analysis versus conceptual analysis

The Referential-Attributive Distinction

Given what we’ve said so far, we have some insight into a problem raised by Keith Donnellan’s work. Donnellan famously distinguished between referential and attributive uses of definite descriptions. Imagine the following. There is some one person – say Jones -- everybody believes to have murdered Smith. I have temporarily forgotten Jones’ name but I wish to make a statement about him -- I recently discovered that he holds a PhD from Yale and I wish to relay this information. So I say

(i) ‘the guy who murdered Smith had a PhD from Yale.’

Under this circumstance, what I wish to communicate, and what I *do* communicate, is a proposition that is *de re* about Jones. Here we have a ‘referential’ use of a definite description.

On the other hand, I can use a definite description to communicate some object-*in*dependent proposition. Suppose nobody knows who murdered Smith; but, given the heinous nature of the crime, it can be reasonably inferred that the murderer, whoever it was, was disturbed. Under this circumstance if I say

(ii) ‘the guy who murdered Smith was insane’,

the proposition I will communicate will be, in effect, *whoever it was that murdered Smith was insane.* And this proposition is not *de re* about Jones (it is about the concept *murderer of Smith*). Here we have an ‘attributive’ use of a definite description.

There can be no denying that Donnellan has made a legitimate distinction here: definite descriptions surely do have both referential and attributive uses. Whether it is a *semantic* distinction is another question altogether. The purpose of this section is to clear up some widespread confusions concerning the attributive use of definite descriptions.

It is generally taken for granted that when a definite description is used attributively, the D-sentence in question is then to be given a Russellian interpretation; i.e. if, in a token of a sentence ‘….the phi…’ , ‘the phi’ is operating attributively, then that sentence is to be read ‘for some x given any y, y=x iff y has phi and…x…’

This is simply not the case. Attributive uses of definite descriptions are no more to be understood on the model of the theory of descriptions than are referential uses. (ii) no more communicates an existence claim than (i) – notwithstanding that (ii) is de dicto, whereas (i) is de re, about Smith.

Whether a definite description is interpreted referentially or attributively is a function of people’s background beliefs. Roughly, if everybody who knows who Smith’s murderer is, then ‘the guy who murdered Smith’ will function referentially; if it is not known who Smith’s murderer is, then ‘the guy who murdered Smith’ will function attributively. Now suppose I don’t know *whether* Smith was murdered or, therefore, how many people murdered Smith; and suppose I apprise my interlocutor of these lacunae in my

knowledge. Under that circumstance, surely, if my interlocutor were to use the expression ‘the guy who murdered Smith’, I could not but interpret that usage attributively. So for example if I say ‘I don’t know whether Smith was murdered or therefore how many people murdered him, supposing that, in fact, he was murdered; and I also, obviously, don’t know anything about the mental state of his murderer or murderers, supposing (once again) that he was in fact murdered’ Under these circumstances, if my interlocutor responds with a sentence of the form ‘…the guy who murdered Smith…’ that sentence will be meant, and interpreted, attributively. But, demonstrably, such a sentence would *not* be given a Russellian reading; in other words, it would not communicate a proposition of the form ‘for some x, given any y, y=x iff y is a murderer of Smith and…x…’: This becomes evident when we consider the following two dialogues, which are exactly alike *except* that, where the latter uses a D-sentence, the former uses the Russellian paraphrase of that sentence (assume that all contextual factors are held constant between the two dialogues):

Dialogue 1:

Me: I know Smith is dead. But I don’t know whether Smith was murdered. Therefore, I don’t know how many people murdered him, supposing that he was in fact murdered. And, needless to say, I don’t know anything about the mental state of his murderer or murderers -- supposing, once again, that he was in fact murdered. I have heard rumors that exactly one person murdered Smith. Are these rumors true?

You: Yes. Exactly one person murdered Smith, and that person was insane.

Dialogue 2

Me: [Same as in previous dialogue.]
You: Yes. *Smith's murderer was insane.

Your response in the second dialogue is deviant because you failed to *assert* exactly how many people murdered Smith. But if Russell's theory applied to your utterance of "Smith's murderer was insane", you wouldn't be guilty of this failing. Of course, your use of that sentence is attributive, not referential. Given any attributive use of a D-sentence, an exactly similar argument could given to show that it was not to be given a Russellian reading. To sum up, even when a D-sentence is functioning attributively, it is not to be given a Russellian reading.

Let us consider a possible objection to our reasoning:

Suppose you are right and that, in the second dialogue, the definite description is not functioning as a quantifier. Now surely if the definite description is not operating as a quantifier, then it must be functioning as a referring expression. But, by your admission, the second dialogue is a context where the definite description is operating attributively *as opposed to referentially.* So by your own admission the definite description is *not* operating referentially.

I have not contradicted myself, though I can see how I might be seen as having done so. So-called 'attributive' uses of definite descriptions are, both semantically *and* pragmatically, referring. If you say 'the guy who murdered Smith is insane', you have *referred to the murderer*, regardless of what

anyone's background knowledge is. Now if the one thing people know about the guy who murdered of Smith is that he is the guy who murdered Smith, then inevitably statements like 'the guy who murdered Smith is insane' will provide *information* only about some individual *qua* satisfier of a certain description: i.e. under that circumstances, such statements will give one information only about a certain description or concept, notwithstanding that *semantically* the definite description attributes a certain property to a certain individual. If the only access you have to a certain individual x is through your knowledge that it satisfies some description – e.g. 'guy who murdered Smith' – then, inevitably, the only thing you can *do* with the information provided by a sentence like 'the guy who murdered Smith is insane' is to apply it towards a description. For under that circumstance you are collaterally blocked, by your ignorance, from applying the predicate in question (*is insane*) to the referent except *by way of* a certain concept (*murderer of Smith*) to a certain description.

But the dynamics in question have nothing to do with semantics; it has to do with your ignorance and your consequent inability to work with the semantics of a particular sentence. Obviously the *semantics* of 'the murderer of Smith' do not change according to how much background knowledge a given auditor has: background knowledge is a quintessentially pragmatic factor. What background knowledge *does* change is how people *access* the proposition semantically encoded in a given sentence (or sentence-token). How people *access* the proposition semantically encoded in a sentence is obviously a function of their epistemic situations (i.e. of what they know and how they know it); and, obviously, gross differences in auditor-knowledge will express themselves in gross differences in what the auditor *does* with semantically encoded information – in, for example, what conclusions he draws from such information. But what the auditor *does* with semantically

encoded information is a paradigmatically *pragmatic* phenomenon.

An analogy might help. Given what Kripke said about names – viz. that, at the level of semantics, they do not have Fregean senses – it seems *de rigueur* to suppose that a name's one semantic value is its referent. At the same time, it is obvious that 'Hesperus is lovely' and 'Phosphorous is lovely' can *communicate* different pieces of information. But this doesn't mean that there is some flaw in Kripke's arguments – it doesn't mean that direct reference theory is wrong. It corresponds only to the banal fact that an object-involving proposition – in this case, that proposition consisting in the ascription of loveliness to <Venus> (the braces are meant to indicate that it is the object *per se*, as opposed to some concept thereof, that is a constituent of the proposition in question) –can be accessed and, once accessed, used in very different ways; that, in other words, the *dynamics* to which a semantically encoded proposition are subject can vary a great deal from situation to situation. But these dynamics are, undeniably, a concern of pragmatics, not of semantics. Semantics is, by definition, concerned with literal meaning *in isolation* of these dynamics; while pragmatics, by definition, is concerned with the dynamics themselves.

The predicative-referential distinction

The considerations just stated enable us to deal with a rather delicate problem. I have argued definite descriptions are referring expressions. But are they *always* referring expressions? Even Strawson, the greatest critic of Russell's theory, said that definite descriptions have a non-referring use. According to Strawson, when you say "Napoleon is the greatest military leader of all time", the definite description is not being used to refer to anyone; it is being used to express a certain predicate, which is to apply to the

subject of the sentence. The definite description is thus being used "predicatively". [54] [55]

Here we must be very sensitive to the distinction between pragmatics and semantics. First of all, there can be no doubt that definite descriptions can be *used* to express predicates. When you say "he is the smartest guy I've ever known", the definite description is surely being *used* predicatively. But given only that an expression can be *used* in different ways, it doesn't follow that it is semantically ambiguous. (The sentence "I can't guarantee that you have a future at this company" can be used to say: *you're fired.* Or it can be used to say: *you're not fired, but there is a reasonable chance you won't make partner*.[JK2])

At the same time, it is very unlikely that "the smartest guy I've ever known" is actually *semantically ambiguous* between referential and predicate meanings. In general, it is unlikely that definite descriptions are semantically ambiguous between predicative and referential meanings. The reason for this lies in a variant of a point made by Kripke.[56] If "the phi" were semantically ambiguous between referential and predicative meanings, then we'd expect there to be languages that had *different* expressions for those two meanings – just as there are languages in which there is one word for bank (financial institution) and another for bank (river's edge). So we'd expect there to be some language L such that the L-translation of "the smartest guy I've ever known" (referential use) would be different from the L-translation of "the smartest guy I've ever known" (predicative use). But, as a matter of empirical fact, there are no such languages.[57] So the referential-predicative distinction that Strawson makes is probably not a *semantic* distinction; it doesn't correspond to any semantic *ambiguity* on the part of definite descriptions. Thus "the smartest guy I've ever known" has the same

semantics in

(*) “The smartest guy I’ve ever known drank like a fish”

and also in

(**) “Bob Jones is the smartest guy I’ve ever known.”

We’ve established that, in (*), the definite description is a referring expression. So the definite description in (**) must also be a referring expression, at least at the level of semantics. How is this to be reconciled with the fact that, in (**), the definite description is being used to express a property, not to denote an individual?

There is no problem here. Remember what we said about “Hesperus” and “Phosphorous”. We know from Kripke that, at the level of semantics (literal meaning), these words have no descriptive content; they are just labels. And they label the same thing. So they have the same semantic value – the same literal meaning. At the same time, “Hesperus is Phosphorous” is obviously non-trivia[JK3]l. This is reconciled with Kripke’s discovery by pointing out that, at the level of pragmatics (not semantics), these terms are associated with descriptive information, and that the descriptive information associated with “Hesperus” is different from that associated with “Phosphorous.”

Now since “Hesperus” and “Phosphorous” are associated (pragmatically, not semantically) with descriptive information, it follows that when you say “Hesperus is Phosphorous” you are, in effect, using the terms “Hesperus” and “Phosphorous” in a *predicative* manner; those terms are being used to *predicate* the descriptions with which they are pragmatically related. That, of course, is why the sentence “Hesperus is identical with Phosphorous” can be

used to communicate something non-trivial, even though what it semantically encodes *is* trivial.

The bottom line is that even *proper names* – which are paradigm cases of referring-expressions, and which we know from Kripke to have no descriptive content for their semantics – can be used predicatively. So given only that definite descriptions can be used predicatively, it doesn't even *prima facie* follow that they are *semantically* predicative.

The following might be said in response to this:

There is an obvious difference between "the smartest guy I've ever known", on the one hand, and "Hesperus", on the other. In general, there is an obvious difference between definite descriptions and names. The descriptive information associated with "the smartest guy I've ever known" is *obviously* not some pragmatic epiphenomenon; it is obviously an essential feature of that expression's semantics. By contrast, the descriptive information associated with "Hesperus" is *not* part of the semantics of that expression, and is just a pragmatic epiphenomenon. Given this, it is hard to believe that, when one uses a definite descriptive predicatively, that usage doesn't have a *semantic* basis in the meaning of that definite description.

It is absolutely true that descriptive content is part of the semantics of "the smartest guy I've ever known." But it does *not* follow, and it is not the case, that that descriptive information is part of the *proposition* semantically encoded in sentences of the form "…the smartest guy I've ever known…" To think otherwise would be to overlook the crucial semantic insights that David Kaplan had in connection with demonstratives.

There are[58], Kaplan observes, two dimensions to the semantics of a demonstrative: its "character" and its "content." Consider the word "I".

Different tokens of this word refer to different people on different occasions. When Frank says “I am tired”, the “I” refers to Frank; when Jerry tokens that same sentence, the “I” refers to Jerry. The person referred to by a given token of “I” is the *content* of that token.

Obviously there is a general rule which assigns particular referents to particular tokens of “I”. That rule is something like: *a token of “I” refers to person who produced that token.* That rule is the “character” of the word “I”.

Notice that the character of “I” is associated with a *definite description*: “the speaker of these words”, or something of the sort. So the character of a demonstrative is identical with, or given by, a description.

All demonstrative expressions have both content and character. Particular tokens of them refer to particular things (so they have content); and there are general rules pairing off specific occurrences of demonstratives with specific things (so they have character).

The *character* of a demonstrative serves merely to *determine* the content of a sentence-token containing that demonstrative, but the character is not itself a part of that content. Suppose Bob says “I am tired”, and let (*) be the proposition that is semantically encoded in Bob’s utterance. For (*) to be true, it is necessary and sufficient that the person who is *in fact* the referent of that token of “I” – namely *Bob* -- have the property of being tired. So (*) is simply *<Bob> is tired*.

Thus the descriptive information associated with “I” is not a part of (*). For (*) is *not* the proposition

(**) *exactly one person is speaking these words, and any such person is tired.*

After all, (**) could be true even if Bob was *not* tired – even if, in fact, Bob didn’t exist. But when Bob says “I am tired”, what he is saying couldn’t

possibly be true unless Bob specifically was tired.

Clearly “I” is a referring term. At the same time, “I” is associated with descriptive content; and this association is obviously semantic.[59] So Kaplan’s work on demonstratives makes it clear how genuinely referential terms can be semantically associated with descriptive content.

Even though demonstratives are referring terms, they can be *used* predicatively. Imagine the following. Your name is John Smith. You are a distinguished professor, so people routinely try to ingratiate themselves to you. You are at a dinner party and somebody comes up to and says: “Do you know which person here is John Smith?” In response you say “I am John Smith.”

For the moment, let us leave aside the question of what is *semantically* encoded in your words; let us focus on what you are *conveying*. You are conveying that John Smith is the person who satisfies a certain *description* – the description *the unique person uttering these very words* or something of the sort. The description in question is the same description that is associated with the *character* of the word “I”. And that, of course, is why you *conveyed* the proposition *John Smith is the person who satisfies the description: the unique speaker of these very words*. In that case, the word “I” is being used predicatively; and the description predicated is the same description that figures in the character of “I”.

Surely we don’t want to say that “I” is *semantically* ambiguous between referential and predicative meanings. We don’t want to say that the word “I” has one semantics in “I am tired” and a different semantics in “I am John Smith.” An analogue of Kripke’s argument shows this: there is no language L such that the L-translation of “I” (referential use: as in “I am tired”) is different from the L-translation of “I” (predicative use: as in “I am John

Smith”).[60]

What is going on here is much less radical than that. The *literal* meaning of a token of a referring term is simply the referent of that token. The literal meaning of “Hesperus” is Hesperus (Venus); the literal meaning of a token of “I” is the person to whom that token refers. But, of course, descriptive information can be associated with referring terms. And referring terms can be *used* to express that descriptive information. “Hesperus” can be *used* to express the predicate “last celestial body to disappear from the morning sky.” And “I” can be *used* to express the predicate “the person uttering these very words.” When a token of “I” is used in this way – when it is used predicatively – that corresponds only to the banal fact that expressions can be used to convey content other than their literal meanings.

Everything we’ve just said about demonstratives applies in spades to definite descriptions, and points to a very definite analysis of the latter.[61] That analysis is as follows. Definite descriptions are referring terms. Like demonstratives, definite descriptions are context sensitive. So “the President of the U.S.A.” refers to different people in different contexts; in 1987, it referred to Ronald Reagan; in 1997, it referred to Bill Clinton. So it is *tokens* of definite descriptions that refer, not the definite descriptions *per se.*

But even though they (or, more precisely, tokens of them) are *referring* expressions, definite descriptions are obviously associated with descriptive information. This descriptive information serves to *determine* what propositions are semantically encoded in sentence-tokens containing those descriptions. But this descriptive information is not itself a part of those propositions.

The referent of a particular token of a referring expression could be thought of as its “content”. And the rule that assigns a referent to different tokens of a

definite description could be thought of as its "character."

So "the guy over there" is a referring term. Thus the semantic content of a particular *token* of "the guy over there" is some particular guy. If Joe Bloggs is the guy who is standing over there, then Joe Bloggs is himself the semantic content of that token of "the guy who is standing there." Thus a token of "the guy who is standing there is lazy", when Joe Bloggs is the guy who is standing in the relevant location, encodes the proposition *<Joe Bloggs> is lazy.*

The descriptive information associated with the definite description serves to *determine* which proposition is semantically encoded in that sentence-token, but it does not figure in that proposition itself. Again, suppose Joe Bloggs is the guy standing over there, and you say "the guy standing over there is lazy." In order for the proposition semantically encoded in that sentence-token to be true, Joe Bloggs specifically must be lazy. If someone else had been standing there, you would have predicated laziness of that other person, and the proposition semantically encoded in your words would have been different. The proposition you've tokened is true iff Joe Bloggs – not anyone else – is lazy.

But the story is different with the proposition *exactly one person is standing over there and that person is lazy.* That proposition could well be true even if someone other than Joe Bloggs is the person standing in the location in question – even if Joe Bloggs had never existed.

Now we can close the argument. We've seen that, even though demonstratives are paradigm cases of referring expressions, they can be *used* predicatively. Exactly the same thing is true of definite descriptions. In the sentence "Napoleon was the greatest military leader of all time", the definite description is *at the level of semantics* operating referentially. But it is being *used* predicatively – just as "I" is being *used* predicatively in the token of "I

am John Smith" that we discussed a moment ago, even though, at the level of semantics, the "I" is operating referentially.

It is no great mystery that definite descriptions – like demonstratives and even proper names – are sometimes *used* predicatively, even though *semantically* they are referring expressions. Grice made it clear that what is communicated by a sentence-token is the result of an interaction between literal meaning and context, plus various expectations and background understandings that the auditor brings with him. One such expectation is that people do not try to communicate trivialities. If somebody says "boys will be boys", what the person is *literally* saying is trivial. The auditor assumes that the speaker doesn't wish to communicate a triviality. The auditor takes the speaker to be *communicating* the most proximal appropriate (non-trivial) message (which is something along the lines of *by nature boys are wild creatures, and it isn't reasonable to expect boys to behave in any other way*). At the semantic level, the sentence token "I am John Smith" is (if true) completely trivial: it is just two referring terms flanking an identity sign. So someone hearing that token will not take the speaker to be communicating that message, and will look for the nearest non-literal message. Inevitably, the auditor pays special attention to the fact that "I" is associated with descriptive information, and that knowledge factors into the proposition which the auditor takes the speaker to be communicating. The same thing is true *mutatis mutandis* of "Napoleon was the greatest military leader of all time" and, indeed, of all predicative uses of definite descriptions.

Conceptual versus Semantic Analysis

There is a rather large concern that I haven't addressed yet:

Throughout this paper, you have presupposed that literal meaning is always transparent, that no real work is needed to figure out what a sentence literally means. Admittedly, your points about implicature constitute a kind of argument for this view. But there must be something wrong with this argument; for it is easy to show that the conclusion of that argument is false: it is easy to show literal meaning is often very hard to ascertain. It was millennia before anyone had any idea how to analyze statements like '1+1=2'; statements about continuity and (therewith) about area. Adequate analyses of concepts like knowledge and justice have yet to be given. Yet you seem to think that it is so easy to grasp: and your argument against Russell's theory obviously fails because you are wrong about this.[62]

This point embodies a crucial confusion. The objector wrongly identifies *semantic* analysis with *conceptual* analysis. Consider the sentence (a) 'that is a sphere'. This admits of very different, but equally correct, analyses, e.g. (b) *that is a closed three-dimensional figure of uniform curvature* and (c) *that is a three-dimensional figure such that, for some point P and some distance D, every surface point on that figure is D from P*. These are the more obvious analyses of (a). But there are others.[1] (In fact, it is provable that there are *infinitely* many other, equally correct analyses.) Now (b) provides a perfectly good analysis of what is encoded in 'that is a sphere', and (c) provides a perfectly good, but different, analysis of what is encoded in that same sentence. But the sentence in question is obviously not *ambiguous*. But if *both* (b) and (c) provided *semantic* analyses of that sentence, then obviously

it would be ambiguous; for (b) and (c), though necessarily equivalent, are plainly not the same proposition. (Actually, if each conceptual analysis of 'that is a sphere' were one of its *meanings* – i.e. were semantically encoded by that sentence – then 'that is a sphere' would have as many meanings as there are axiomatizations of geometry: that is, infinitely many.)

Of course, one might respond by saying that (b) and (c) are, in fact, the *same* proposition; that the sentences that express them are synonymous. But this is patently not the case. Consider the sentence:

(d) 'that is a closed three-dimensional figure of uniform curvature'

and

(e) 'that is a three-dimensional figure such that, for some point P and some distance D, every surface point on that figure is D from P.'

The propositions encoded in these are composed of very different concepts; one involves the concept *point*; the other does not. One involves the concept *curvature*; the other does not. And, what is connected with this, (e) is not a mere rephrasing of (d); and neither is (d) a mere rephrasing of (a). What (e) semantically encodes consists of totally different concepts from what (d) encodes, and is therefore a different entity altogether. Given this, it would make nonsense of the concept of synonymy to say that (d) and (e) are synonymous.

We can use our synonymy test to establish the non-synonymy of (d) and (e), and also the non-synonymy of (a) with either of those two. Consider the following two dialogues:

Dialogue 1

A: That object over there is a three-dimensional figure such that, for some point P and some distance D, every surface point on that figure is D from P.
B: Ok, but I still don't know whether that object has uniform curvature.

Dialogue 2

A: That object over there is a closed three-dimensional figure of uniform curvature.
B: *Ok, but I still don't know whether that object is of uniform curvature.

In the first dialogue, B's response has shown a lack of *mathematical* acumen. But *linguistically* B's response is perfectly fine. We cannot accuse B of having failed to take in what A has said; we can only accuse him of not having the mathematical virtuosity to draw some rather subtle inferences from what A has said.

In the second dialogue matters are different. A has directly *told* B that the object in question has a uniform curvature. B's response shows a failure to take in what A has just *told* him. B's response embodies *not* a failure to derive the mathematical *consequences* of what A has told him, but merely to register the very thing that A has told him.

Neither dialogue contains a sentence with any tone. And the difference between them – that B's response in the first is appropriate, while his response in the second is radically defective – obviously doesn't have to do with context. (Even if both dialogues occur in qualitatively identical contexts, B's response in the second dialogue will be deeply perverse, while his response in the first will be appropriate.) Finally, the difference obviously

cannot be chalked up to *incidental* information. For the problem in the second dialogue obviously lies in B's speech-act itself, not in some incidental by product of it.

The only relevant difference between the two dialogues is that, in one of them, A uses (d) while in the other A uses (e). Everything else is the same. We must conclude, then, that the proposition semantically encoded in A's words differs between the two dialogues: (d) and (e) encode different propositions.

An exactly analogous argument could be constructed to show that "that is a sphere" encodes a different proposition from either (d) or (e). To sum up, *distinct* proposition can be logically equivalent; conceptual analysis provides logical equivalences, not identities of meaning. So conceptual analysis is not semantic analysis.

Why the Theory of Descriptions Cannot be Construed as a Piece of Conceptual Analysis

This enables us to dispose of another possible objection to what we've said. Perhaps, one might argue, Russell *is* providing a conceptual analysis of D-sentences: he isn't giving us their *semantics* or meanings. Rather, he is giving us conceptually necessary and sufficient conditions for the *truth* of such sentences – just as somebody who analyses numbers as sets of equipollent sets is giving us a *conceptual* analysis of number-sentences, not a semantic or linguistic analysis of them.

I will leave aside the historical question of how Russell meant his theory to be taken. What is clear is that *if* Russell's theory is taken as a piece of conceptual analysis, and not as a piece of meaning-giving, then it is quite

obviously a failure.

The essence of Russell's theory is that "the" (in the singular) is a special kind of quantifier. And that quantifier is contextually defined thus:

(a) "…the phi…"

is true iff

(b) *there is exactly one phi and…that thing…*

For the sake of argument, suppose that Russell's theory does not give the *meaning* of "the"; in other words, let us suppose that Russell is not doing a *semantic* analysis of "the" at all. In that case, the meaning – the semantic content – of "the" is *not* the quantifier just defined. So what would the meaning, the semantic content, of "the" be? It seems quite clear that *if* the semantic content of "the" is any kind of quantifier, it must be the one Russell has in mind. (In other words, it must be some quantifier that is defined in more or less the following way: "…the phi…" is true iff *exactly one thing has phi and…that thing…* or *for some x, given any y, y=x iff y has phi, and… x…*) Surely Russell is right to *this* extent: *if* "the" is some kind of quantifier, it is definitely the quantifier he has in mind, or one not essentially different from it.

Now we've made a case that "the" is *not* the quantifier Russell has in mind. Given what we just said, it follows that "the" is not *any* kind of quantifier. Thus "the man", "the Queen of England" and definite descriptions generally, are not quantifiers, at least not at the level of semantics.

Given that, at the level of semantics, such expressions are not quantifiers, what could they be? There is really only one option: they must be referring

expressions. When we consider the class of all those things that grammarians (not philosophers) refer to as “nouns” and “noun-phrases”– e.g. “Socrates”, “Mortimer Adler”, “a man”, “two penguins” – they are all, at the level of semantics, either referring expressions *or* they are quantifiers. Given that “the man over there” is semantically not a quantifier, it must semantically be a referring expression.

But if “the man over there” is *semantically* a referring expression, then Russell’s analysis – even if it is construed as a piece of *conceptual* analysis, as opposed to semantic analysis – is dramatically wrong. If “the guy over there” is semantically a referring expression, then “the guy over there is rich” encodes a *singular* proposition – one that has the form:

(*) *<x> is rich.*

Of course, x is an *individual.* So “the guy over there is bald” encodes a proposition of the same form as *Frank is rich.* Now *no* proposition of the form *<x> is rich* entails, or is entailed by, by anything of the form

(**) *for some x, given any y, y=x iff y has phi, and x is rich.*

(**) and (*) simply *don’t* have the same truth-conditions. (Obviously the proposition *Frank is rich* doesn’t entail that for some property phi, exactly one person has phi, and that person is rich. Nor does the converse entailment hold.) *A fortiori* (*) and (**) don’t have *necessarily* the same truth-conditions. So (**) is *not* a conceptual analysis of (*). We’ve also seen, of course, that Russell’s theory doesn’t give a correct semantic analysis of “the” in singular. So Russell’s theory is a failure *tout court.*

How synonymous sentences can differ in sense (cognitive value)

I'd like to end this paper by considering an important problem that our analysis faces. I have said that a token of "the man over there is rich" semantically encodes a proposition consisting of a certain individual and the property of being rich, and thus semantically encodes the same proposition as "Jack is rich", assuming that Jack is the man over there. But one can believe that Jack is rich without believing that the man over there is rich, and *vice versa*. Thus, it would seem, "the man over there is rich" (or, more accurately, tokens of that sentence) cannot possibly encode the same proposition as "Jack is rich." How are we to deal with this problem?[63]

First of all, we must recognize that this problem does not have to do specifically with definite descriptions. The same problem characterizes sentences containing demonstratives and also sentences containing proper names; and there is no doubt that tokens of such sentences encode singular (object-involving) propositions. It is undeniable that tokens of "he is rich" and "you are rich" semantically encode singular propositions: propositions consisting of an individual and a property. But as Kaplan points out, I can believe what is expressed by one demonstrative sentence-token and disbelieve what is expressed some other demonstrative sentence-token, even if both tokens semantically encode the same proposition. To use Kaplan's own example, I can believe what is expressed by "*that* man's pants are on fire" and not believe what is expressed by "my pants are on fire" -- even if I am that man.[64] But there is no doubt that, if the demonstrative expressions in those sentence-tokens are co-referential, then both sentence-tokens semantically encode the same proposition.

And of course there are the familiar Frege cases: I can believe what is

expressed by “Hesperus is lovely” without believing what is expressed by “Phosphorous is lovely”, even though – as we now know from Kripke – both sentences semantically encode the same singular proposition: one consisting wholly of a planet and the property of loveliness. To sum up, it is well known that sentences (or sentence-tokens) that semantically encode the same proposition can differ in cognitive value.

But what is the explanation of this fact? In effect, we have already seen what the explanation is. Consider some token of “he is rich” and also of “you are rich”, and suppose that the demonstratives are co-referential. Let (*) and (**) be these two sentence-tokens. Remember what Kaplan says about demonstratives: any demonstrative is associated with descriptive information (“character”), but this information does not figure in the propositions semantically encoded in sentence-tokens containing that demonstrative. Rather, that descriptive information serves merely to *pick out* the appropriate object and, therewith, the proposition semantically encoded in the relevant sentence-tokens. So, yes, (*) and (**) do differ in cognitive value; and this is because they are associated with different descriptive information. But this descriptive information is not part of the proposition semantically encoded in either (*) or (**). So the cognitive difference – the difference in “sense” – between (*) and (**) has to do not with the proposition that they encode, but with the manner in which they *pick out* that proposition: they pick out the *same* proposition in different ways.

Obviously exactly similar remarks apply to “the guy over there is rich” and “Jack is rich”. Both sentence-tokens encode the same proposition. But they differ in cognitive value; for they are associated with different descriptive information. This descriptive information is not part of the proposition that they semantically encodes; it serves merely to *pick out* that proposition.

To sum up, the difference in cognitive value or sense between “the man over there is rich” and “Jack is rich” in no way threatens a referentialist account of definite descriptions.

Further, it is clear how this analysis explains the difference in cognitive value between “Hesperus is lovely” and “Phosphorous is lovely.” The two sentences are associated with different ways of picking out the same singular proposition. So they are associated with different descriptive information. But this descriptive information serves merely to pick out a certain proposition – the same in both cases – and it does not figure *in* that proposition.

Of course, we know from Kripke that *at the level of semantics* “Hesperus” and “Phosphorous” are not associated with descriptive information; those expressions are merely labels. Given this, some might balk at what we said in the last paragraph. But this would not be the right reaction. For clearly *at the level of pragmatics* names can be associated with descriptive information: indeed, they must be. When you learn what a name refers to, you learn it *not* in a vacuum, but in connection with some particular presentation or description. Someone can tell you what “Hesperus” means only by describing what it refers to or by pointing to it (or perhaps through some mixture of the two). In either case, it is *through* descriptive information that you learn what “Hesperus” refers to. If somebody *describes* what “Hesperus” refers to, they must *explicitly* use some definite description: they must say “it is *the second planet from the sun*” or something of the sort. If somebody *points* to Hesperus, and says “that is what ‘Hesperus’ refers to” (or simply “that is Hesperus”), it is, as we just noted, through *descriptive* information that you learn what that word refers to; for any ostension is packed with descriptive information. This is because a presentation, and thus an ostension, of Hesperus is not *just* a presentation of Hesperus: it is a presentation of an object that has a certain appearance and a certain shape and a certain location

in the evening sky. Things cannot just be presented; they must be presented *as* having these or those characteristics. So if you learn the meaning of a word through a presentation or ostension, you cannot but associate a great deal of descriptive information with that word. (And if somebody learns the meaning of that word through a qualitatively *different* presentation, that person will associate *different* descriptive information with that same word.) So even though Kripke is quite right to say that, at the level of semantics, "Hesperus" and "Phosphorous" are merely labels, not descriptions, it doesn't follow – and it isn't the case – that *at some other level* we don't associate any descriptive information with these terms. And obviously sentences of the form "…Hesperus…" differ in cognitive value from sentences of the form "….Phosphorous…" precisely because we *do* associate different descriptive information with "Hesperus" and "Phosphorous." (Kripke showed that this association is pragmatic, not semantic, in nature; Kripke did *not* show that there is *no* such association.) The one word is associated with a presentation of an object seen at night, the other with a presentation of an object seen in the morning. And it is plainly this presentational difference – this difference in descriptive information – that accounts for the cognitive difference between "…Hesperus…" and "…Phosphorous…"

But – to close the present discussion -- it is also clear that "… Hesperus…" and "…Phosphorous…" encode the same singular proposition. Thus the aforementioned descriptive information accounts for the difference in cognitive value between those two sentences *not* by figuring *in* the proposition that they semantically encode, but rather by *picking out* that one proposition in different ways.

We can obviously know object-dependent truths. If I see a dog chasing a cat, I am aware of an object-dependent truth: one consisting of a certain cat and a certain dog and the relation of chasing. But the dog and the cat cannot

just be presented to me; they must presented to me *as* having these or those properties. As Barry Stroud says, all perception is "predicational" [65]: when I see the dog I must see it *as* having these or these properties; so I must, in effect, predicate these or these properties of the dog. Thus, when I become aware of some object- dependent truth like *the cat is chasing the dog*, it is necessarily *by way of* predicational or descriptive information. By the same token, I can be aware of the *same* proposition through *different* descriptive information. (To switch examples, I can become aware of the fact that Venus is lovely by looking at Venus through a telescope, by looking at Venus from a space-ship that is right next to it, by being *on* Venus, and so on. In all these cases, there is some *one* proposition – namely *Venus is lovely* – of whose truth I am becoming aware. But, of course, I am, in each case, become aware of this truth through different descriptive information.) This phenomenon constitutes the epistemological basis for the fact that different sentences which encode one and the same proposition can have very different cognitive values.

Bibliography

Meaning and Truth. (1991.) Jay L. Garfield and Murray Kitely (eds.)
St. Paul, Minnesota: Paragon House

The Philosophy of Language. (1990.) Second Edition A.P. Martinich (ed.).
Oxford: Oxford University Press.

Blackburn, Simon (1984.) *Spreading the Word* Oxford: Clarendon Press.

Dummett, Michael (1981.) *Frege: Philosophy of Language* Second Edition
Cambridge: Harvard University Press.

Graff, Delia. "Descriptions as Predicates" in *Philosophical Studies* 102, 2001, pp.1-42.

Grice, Paul (1989a) 'Logic and Conversation' in *Studies in the Ways of Words* Cambridge: Harvard University Press.
-- (1989b) 'Presupposition and Implicature' in *Studies in the Ways of Words.*
Cambridge: Harvard University Press.

Kaplan, David (1989.) 'Demonstratives' in *Themes from Kaplan.*
Oxford: Oxford University Press

Kripke, Saul. (1991) ‘Speakers Reference and Semantic Reference’. In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.

MacCawley, James D. *Everything that Linguists Have Always Wanted to Know about Logic* [*But were afraid to ask]* Second Edition Chicago: University of Chicago Press 1993.

Neale, Stephen. *Descriptions.* Cambridge: MIT Press 1990.

Russell, Bertrand ‘On Denoting’ in A.P. Martinich (ed) *The Philosophy of Language* Oxford: 1990

Salmon, Nathan. *Frege’s Puzzle.* Ridgeview Press. Atascadero: 1986

Stroud, Barry *The Quest for Reality* Oxford University Press: New York 2000

Strawson, Peter. ‘On Referring’ in A.P. Martinich (ed) *The Philosophy of Language.* Oxford: Oxford University Press.

Non-declarative contexts and the theory of descriptions

This essay will attempt to discredit Russell’s so-called ‘theory of descriptions’ by showing that it provides a faulty analysis of definite

descriptions[66] occurring in non-declarative sentences – specifically, in questions and imperatives.

Let us refer to sentences containing *definite descriptions* as 'D-sentences'. And let us refer to assertions containing definite descriptions as 'D-assertions', to imperatives containing definite descriptions as 'D-imperatives', and so on. Russell's Theory of Descriptions purports to give the *semantics* of such sentences; i.e. it purports to give their *literal meanings.* The Theory of Descriptions is essentially this. Given a sentence of the form:

(i) '….the phi….'

that sentence has its literal meaning

(ii) *There is exactly one object x having phi and…x…*

or, more perspicuously,

(ii_F)[67] *For some x, x has phi and given any y, y has phi iff x=y and…x…*

So, according to Russell's theory, the sentence

(iii) 'The king of France is bald'

has as its literal meaning the proposition

(iv) *There is exactly one king of France and that person is bald.*

or, more perspicuously,

(iv$_F$) *For some x, x is a king of France and for any y, y=x iff y is a king of France, and x is bald.*

In this essay the distinction between sentences and propositions will be important. Given this, I should make a point that might seem trivial, but that will serve us well later on. (iv) is not a *sentence*: it is a *proposition.*[68] But uncontroversially

(v) 'There is exactly one king of France and that person is bald'

is *a* sentence that has (iv) as its literal meaning. So Russell's theory, in effect, predicts that (iii) and (v) are synonymous sentences. Let us say that (v) is a 'Russellian paraphrase' of (iii). So the term 'Russellian paraphrase' will be used to refer to *sentences*, not to propositions. (iv) is not a Russellian paraphrase of (iii) – for (iv) is a proposition, not a sentence. (v) is not the only Russellian paraphrase of (iii). Indeed,

(v$_F$) 'For some x, x is a king of France and for any y, y=x iff y is a king of France, and x is bald'

is a more perspicuous paraphrase of (iii) than is (v), i.e. (v$_F$) presumably gives a more fine-grained representation of the proposition that, if Russell's theory is correct, is the literal meaning of (iii). The distinction between (v) and (v$_F$) – between, in general, the more and the less formal ways of paraphrasing D-sentences – will play an important part in our inquiry: and that is why I am making such heavy weather of this seemingly trivial

distinction.

Though Russell's theory is not as popular as it used to be, it still has no shortage of distinguished proponents[69]; and the merits of this theory continue to be bitterly debated. But the debate is concerned almost exclusively with the question whether D-*assertions* conform to Russell's theory: D-*questions* and *D-imperatives* are seldom discussed.[70]

In this paper I wish to show that Russell's theory does not give the semantics of D-questions and D-imperatives; more specifically, that it does not give the semantics of definite descriptions that occur in questions and imperatives. If this is correct, it follows, I submit, that Russell's theory does not give the semantics of definite descriptions occurring in *any* sentence. Why does this follow? Presumably the expression 'the king of France' has the same semantics in 'the king of France is going to the store' that it has in 'is the king of France going to the store?' and 'make the king of France go to the store'. So *if* indeed Russell's does not provide the semantics for 'the king of France' as it occurs in questions or imperatives containing that expression, then it seems unlikely that Russell's theory provides the semantics for that expression as it occurs in any sentential context.

I will give two arguments for my contention that Russell's theory doesn't provide the right semantics for definite descriptions occurring in questions and imperatives. First, when we look at the behavior of D-questions and D-imperatives – specifically, when we look at what is communicated (asked and ordered) by them – we find that *even after non-semantic factors have been taken into account* these sentences deviate from their Russellian counterparts.

Second, Russell's theory assigns nonsense to questions and imperatives that, quite obviously, make perfectly good sense.

First of all, we must note a fact about D-sentences generally. Whenever a definite description occurs within the scope of any kind of operator, Russell's theory recognizes *two* distinct paraphrases of it. Consider the sentence (Russell's own example):

(*) George IV wanted to know whether Scott was the author of Waverly.

This sentence (as Russell himself stated clearly) is given two different readings by the Theory of Descriptions:

(**) There is some x such that x uniquely wrote Waverly and George IV wanted to know whether x was Scott

and

(***) George IV wanted to know whether exactly one person x wrote Waverly and x was Scott.

(**) gives the definite description 'wide-scope' and (***) gives it 'narrow' scope. All questions (and imperatives) containing definite descriptions are capable of both narrow- and wide-scope readings.

In general, whenever a definite description occurs within the scope of an operator, the sentence is, according to Russell's theory, ambiguous between 'wide-scope' and 'narrow-scope' readings.

Given this, consider the question:

(i) Does your neighbor drink?

As I just pointed out, there are two different Russellian paraphrases of (i). The narrow-scope reading is:

(ii) Is it the case that you have exactly one neighbor and that he/she drinks?

Formally, it is:

(ii_F)[71] Is it the case that there is some x such that, for any y, y=x iff y is a neighbor of yours, and x drinks?

The wide-scope reading of (i) is (very informally):

(iii) You have exactly one neighbor: does that person drink?

Expressed (more) formally, (iii) becomes:

(iii_F) There is some x, such that given any y, y is a neighbor of yours iff y=x, and does x drink?

At this point, two questions must be asked. The first concerns (iii_F): Does this sentence make any sense? Is it well-formed? Uncontroversially, *(iii)* is well-formed. So it might seem that (iii_F) must be well-formed and that, so far, Russell's theory is fairing perfectly well. But this is not the case. For (iii) 'you have exactly one neighbor: does that person drink?' is not *strictly* speaking the Russellian paraphrase of (i) 'does your neighbor drink?' Rather, it is a kind of vernacular gloss of the real Russellian paraphrase. (iii) captures the *macro*-features of the wide-scope Russellian reading of (i). So, in that

attenuated sense, it is a wide-scope Russellian paraphrase (i). But (iii) is not a perspicuous representation of the *micro*-structure of wide-scope Russellian reading of (i). It is *(iii$_F$)* that perspicuously represents this microstructure.

So, granted that (iii) is well-formed, if indeed (iii$_F$) is *ill*-formed, this would be a problem for Russell's theory: it would be a problem because the sentence it paraphrases ('does your neighbor drink?') is obviously well-formed: and with regard to any theory that regards a well-formed, meaningful, sentence as synonymous with an ill-formed, and thus meaningless sentence, any such theory is obviously false.

And (iii$_F$) *does* seem to be ill-formed: at the level of literal meaning, it is *both* a question and an assertion. And it is by no means clear whether a single sentence can, at the level of literal meaning, be both of these things. Obviously a sentence that is *pragmatically* an imperative may be *semantically* an assertion: 'I'd like a vodka tonic', when spoken to a waiter, is semantically declarative and pragmatically imperative. And a sentence that is semantically a question can pragmatically be an assertion. This is the case with rhetorical questions ('Is it so wrong to have compassion for a foe?') But (iii$_F$) is in a completely different category. For (iii$_F$) is *semantically* both a question and an assertion. I will argue (iii$_F$) is indeed ill-formed, the reason being that no well-formed sentence can, at the level of literal meaning, be *both* a question and an assertion.

Now *if* the wide-scope paraphrases of D-questions and D-imperatives are indeed ill-formed, then Russell's theory is in trouble; for if that is the case, then Russell's theory renders 'did the king of France go to the store' capable of a reading whereby it is synonymous with an *ill-formed* sentence. Now an *ill-formed* sentence is one that fails to semantically encode *any* proposition. So, of course, if a sentence S is synonymous with a sentence S' that is ill-formed, and that consequently does not encode *any* proposition, then S itself

fails to encode any proposition, i.e. S itself is non-sensical . But ‘did the king of France go the store?’ seems very much *not* to be synonymous, on any legitimate disambiguation, with an ill-formed sentence; that question seems to be a significant and well-formed on any legitimate reading of it.

The second question is to be understood in terms of the first one. I pointed out that *if* Russell’s theory makes D-questions and D-imperatives be ambiguous between well-formed propositions and nonsense, then Russell’s theory is, to that extent, in quite a bit of trouble. But let us suppose, in deference to Russell’s theory, that the wide-scope paraphrases of D-questions and D-imperatives *are* well-formed. Now supposing that they are well-formed, it is clear enough what propositions are to be associated with them. The wide-scope paraphrase of. ‘did the king of France go to the store?’ is ‘for some x, given any y, y=x iff y is a king of France and did x go the store?’ Now *supposing* that sentence to be well-formed, it is clear what proposition would be associated with it, namely: *someone is uniquely a king of France*: *did that person go to the store?* Of course, the narrow-scope paraphrase of ‘did the king of France go to the store?’ is (essentially) ‘is it the case that there is exactly one person x that is a king of France and that went to the store?’ So (assuming that Russellian wide-scope paraphrase of D-questions are well-formed) if Russell’s theory is correct, then when one asks ‘did the king of France go the store?’ one is either asking (inter alia) whether there is exactly one king of France or one is asserting (inter alia) that there is exactly one king of France.

But it surely *seems* that when you ask someone ‘does your neighbor drink’ you are neither asking whether they have exactly one neighbor nor *asserting* that they have exactly one neighbor.

An exactly similar point applies to D-imperatives. The wide-scope of paraphrase of e.g. ‘tell my maid to come on Tuesday’ is ‘for some x, given

any y, y=x iff y is a maid of mind, and tell x to come on Tuesday.' Now *supposing* that this sentence is well-formed, it is clear enough what proposition it encodes, namely: *there is someone who is uniquely my maid*: *tell that person to come on Tuesday.* Of course, the Russellian narrow-scope paraphrase is 'make it the case that I have exactly one maid x and that x come on Tuesday.' So if Russell's theory is correct when one says e.g. 'tell our maid to come on Tuesday' one is either *ordering* that one's objector make it the case that one has exactly one maid' or one is asserting that one has exactly one maid.

But it surely *seems* that when you say 'tell the maid to come on Tuesday' you are neither ordering your interlocutor to make it the case that you have exactly one maid, nor asserting that you have exactly one maid.

In section II, I will argue that, indeed, when one asks 'does your neighbor drink?' one is, at the level of literal meaning, neither asking whether one's auditor has exactly one neighbor nor *asserting* this; and that consequently Russell's theory doesn't capture the semantics of D-questions.

In section III, I will argue that, in fact, Russellian wide-scope readings of D-questions are *ill*-formed: they are semantic nonsense, even though the sentences they paraphrase are plainly not nonsense on any legitimate reading of them.

In section IV, I will show that everything so far said about D-questions applies to D-imperatives.

The argument set forth in this paper might be seen as having the form 'either P or not P': either the wide-scope paraphrases of D-questions and D-imperatives are well-formed or they are not. If they are well-formed, then Russell's theory is false (for, if well-formed, they encode the wrong propositions). If they are not well-formed, then Russell's theory is false (for, if ill-formed, they are, ambiguous between non-sense and significance,

whereas the sentences they paraphrase are not correspondingly ambiguous). Therefore Russell's theory is false.

II. In this section, I will assume that Russellian wide-scope paraphrases of D-questions are well-formed and that, consequently, the wide-scope reading of 'does the king of France drink?' is *there is exactly one king of France: does he drink?* But I'd like to start by making a point about pragmatics that has to do with D-sentences generally, and not just with D-questions.

Consider the sentence

(i) John is searching for the fountain of youth.

From the standpoint of the Theory of Descriptions, this sentence is capable of both a wide-scope and a narrow-scope reading. The wide-scope reading is:

(ii) For some x, x is uniquely a fountain of youth, and John is searching for x.

The narrow-scope reading is a little harder to generate, but there can be no doubt that it is at least roughly this:

(iii) John believes that, for some x, x is a uniquely a fountain of youth, and John is trying to make it the case that, for any fountain of youth y, John find y.

Given this, consider the following sequence of sentences:

(a) "There is no fountain of youth. Therefore the proposition ***John is searching for the fountain of youth*** is false."

If the Theory of Descriptions is correct, the boldfaced part is ambiguous between a proposition that will make (a) express a valid argument and a proposition that will make (a) express an invalid argument. (If the boldfaced part is assigned (ii) as its meaning, the argument will be valid. If it is assigned (iii) as its meaning, the argument will be invalid.) So, if Russell's theory is right, then (a) as a is *ambiguous* between a good argument and a bad argument.

But obviously (a) is unambiguously a *bad* argument. It is a paradigm case of a bad argument; no logic teacher would ever give it a passing grade. Its unambiguously bad nature becomes clear when it is contrasted with a sequence of sentences that is clearly ambiguous between good and bad[72] arguments.

(b) "Bob is a compulsive gambler. That is why **Bob cannot stop gambling**."

The boldfaced part is ambiguous between (c) *Bob cannot cease to gamble* and (d) *Bob cannot prevent people in general from gambling.* Someone reading (b) would inevitably assign meaning (c) to the boldfaced part: for (b) makes a lot of sense if read in this way. And one would not assign meaning (d) to the boldfaced part: for (b) makes little sense if read in *this* way. Basically, when someone competent is confronted with a sentence S that is ambiguous between two propositions, one of which does the fit the context while the other does not, one strongly tends to give S the reading that will make it best fit the context.

To take another example, suppose someone says:

(e) "**Visiting relatives can be a nuisance**.[73] That is why I don't allow my relatives to visit me."

The boldfaced part is ambiguous between (f) *it can be a nuisance for one to visit one's relatives* and (g) *it can be a nuisance for one to receive visitors who are one's relatives.* If the boldfaced part is assigned meaning (g), then (e) makes sense. If the boldfaced part is assigned meaning (f), (e) makes no sense. Not surprisingly, one would have a strong inclination to assign meaning (g) to the boldfaced part.[74]

To sum up, if a sentence-token S is ambiguous between a proposition P1 that would be wildly out of context, and a proposition P2 that would fit context well, we tend to assign proposition P2 to S.[75] Let us refer to this as the *principle of charitable disambiguation* (PCD).

A corollary of PCD is this. If we encounter a sentence-token S in a certain context C, and we have absolutely *no* tendency to assign a meaning to S that would make it fit C, that means that S does not have that meaning.

In addition to being an argument in its own right against the theory of descriptions, this line of thought gives us a way to show that Russell's theory gives the wrong analysis of questions and imperatives. Consider the question.

(i) Does your neighbor drink?'

According to Russell's theory, this is ambiguous between meanings that are at least *roughly* (that is, in terms of their macro-structures)

(ii) *You have exactly one neighbor: does that person drink?*

and

(iii) *Is it the case that you have exactly one neighbor, and that person drinks?*

Now consider the following dialogue (suppose that B is a well-known billionaire, and that A is a nosy sociologist who is trying to prove a strange theory):

A: I'm a sociologist, and I'm doing research on people's drinking habits. More specifically, I'm interested in the drinking habits of neighbors of billionaires. I believe that any neighbor of a billionaire is a heavy drinker.

B: My dear Sir. Permit me to interrupt you right there. I have a rather peculiar form of amnesia. I have forgotten how many neighbors I have. But once you give me this information, I will remember everything about them, and I will be able to answer all your questions about any neighbors I might have.

A: ***Does your neighbor drink?**[76]

The first point to make is that A's response to B is decidedly odd. But given PCD, if the Theory of Descriptions were correct, A's response *wouldn't* be odd. According to theory of descriptions, 'does your neighbor drink?' has as one of its meanings *you have exactly one neighbor: does that person drink?* In the context in question, *if* 'does your neighbor drink?' could be read as meaning 'you have exactly one neighbor: does that person drink?', it *would* be read that way. For reading it that *would* make A's response fit the context. But no competent speaker of English does read A's response in a

way that makes it fit the context.

A similar argument shows that the Russellian wide-scope paraphrase of 'does your neighbor drink?' is probably not among the meanings actually borne by that sentence. Consider the following dialogue (again suppose that A is a nosy sociologist with a strange theory):

A: I know you to be an intelligent person, and I believe that all intelligent people have exactly one neighbor and that he/she drinks. But I don't know if you have exactly one neighbor.

B: Unfortunately, because of illness, I never leave my house. So I don't know how many neighbors I have.

A: Neither do I. So if it isn't too much trouble, I'd like you to find the answer to the following question: ***Does your neighbor drink?**

First of all, in the context in question, the boldfaced part sounds very strange; it is obviously not appropriate of A to use that sentence here. (For it sounds as though he is *presupposing* that B has exactly one neighbor when, by his own admission, this is just what he wants to know.) But if Russell's theory were correct, then given PCD, the boldfaced sentence would be appropriate. But it isn't appropriate; so Russell's theory is wrong.

III. In this section, I will set forth another argument to the effect that Russell's theory does not provide the right semantics for D-questions. But some review of what I've said so far may be necessary to put this argument in its proper dialectical context.

So far we have operated on the assumption that the wide-scope

paraphrases of D-questions are well-formed; and we have been assigning them the meanings that *if they were* well-formed, they would bear. To make this point concrete: the wide-scope paraphrase of (e.g.) 'did the king of France go to Spain?' is (essentially)

(*) 'there is exactly one king of France x, and did x go to Spain?'

So far we have assumed that (*) is *not* nonsense. Accordingly, we supposed the meaning of (*) to be that proposition that, *given* the aforementioned assumption, is most plausibly seen as being the meaning of (*). So we have supposed the meaning of (*) to be

(**) *There is exactly one king of France: did that individual go to Spain?*

We have seen that *if we operate* within the framework just described, Russell's theory is not viable, as it mis-predicts what D-questions in fact communicate. In other words, *if* the wide-scope paraphrases of D-questions *are* well-formed, then Russell's theory is false.

On the other hand, if the wide-scope paraphrases of D-questions are *not* well-formed, then Russell's theory is false. There are two reasons for this. First, there is no reason to believe that all D-questions and D-imperatives are, on any legitimate reading of them, synonymous with ill-formed sentences: they unambiguously make good sense.

Second, as we saw in the last section, the narrow-scope paraphrases of D-questions *clearly* do not give the right meanings of those sentences, at least in many contexts. So if the narrow-scope paraphrases give the only well-formed paraphrases of such sentences, then Russell's theory simply doesn't assign the right meanings to those sentences. In other words, if Russell's theory is to

be correct, then the wide-scope paraphrase of D-questions must be well-formed.

But the wide-scope paraphrases of D-questions are *not* well-formed. This is what I will now show.

Let us revisit the facts that formed our point of departure. Consider the sentence:

(i) Does your neighbor drink?

Russell's theory assigns two readings to this question: a wide-scope and a narrow-scope reading. The narrow-scope reading is:

(ii_F) Is it the case that there is some x such that, for any y, y=x iff y is a neighbor of yours, and x drinks?

The wide-scope reading is:

(iii_F) There is some x, such that given any y, y is a neighbor of yours iff y=x, and does x drink?

(ii_F) is unquestionably well-formed; and we will say nothing more about it. But is (iii_F) well-formed? No.

First of all, what is (iii_F)? Is it a question or an assertion? It is some kind of a combination of both. (But it is not a *conjunction* of an assertion and a question, since the main connective is the existential operator.)

Uncontroversially, a single sentence can *at one level* be an assertion and *at*

another level be a question. Rhetorical questions are semantically questions, while pragmatically they are assertions. But (iii_F) is on quite another level: it is *semantically* both a question and an assertion. Does (iii_F) make sense? [77]

It does not. Consider the sentence

(i) For some x, x is such that x is uniquely a king of France and did x go the store.

This sentence is both assertoric and interrogative. The interrogative part is 'did x go to the store?' But notice that this is *not* really a question; for it contains a free variable. So (i) contains *no* question, and thus doesn't ask *anything*. But according to Russell's theory, (i) is the paraphrase of 'did the king of France go the store?', which obviously *does* ask something. So (i) doesn't capture the meaning of 'did the king of France go the store?', contrary to what Russell's theory predicts.

Before proceeding, I should address an objection:

Admittedly, it would be meaningless to ask 'did x go to the store?' *simpliciter*. But it would not be meaningless to ask 'did x go to the store' *for some value of x*. (A comparison: it would be meaningless to say 'necessarily x is even' *simpliciter*. But it would not be meaningless to say 'x is even' *for some value of x*.) And this appears to be just what (i) is doing. (i) expresses the question '*for some value of x*, did x go to the store?' [78]

Let us be quite clear on what the objector is saying. According to the objection, (i) embodies the question

(ii) ‘for some x, did x go the store?’

First of all, we must note that (ii) is not the same question that is asked by

(iii) ‘is it the case that, for some x, x went to the store?’

(iii) *is* a perfectly good question. But it is not the same question as (ii). The difference is that, in (ii), the existence-operator has a *wider scope* than it has in (iii). (ii) is what results when part of (i) is deleted – when the clause ‘x is such that it is uniquely a king of France’ is removed. Otherwise (ii) is just like (i). Now the existence operator (‘for some x’) in (i) and (ii) has assertoric force; so it does not have interrogative force . So the existence operator in (i), and therefore in (ii), does not fall within the scope of an ‘interrogative operator’ (in any case, this is one way of thinking of the matter). Indeed, the ‘interrogative operator’ in (i) and (ii) falls within the scope of the existence-operator, not *vice versa.* So the existence-operator has a wider scope than the interrogative operator in (i) and (ii).

In (iii), on the other hand, the existence-operator plainly has interrogative force. So the existence operator falls *within* the scope of an interrogative operator: the former has narrower scope than the latter.

Because the existence-operator has wider-scope in (i) and (ii) than the interrogative operator, it follows that (ii) – in so far as it is well-formed and asks *any* question – asks a question that is in effect *de re* about somebody, i.e. that is about some *particular* individual. In other words, (ii) says ‘for some particular x, did x go to the store’. (iii), on the other hand, does not ask that sort of question; it does not ask a question about any *particular* person.

Now that we’ve taken care of these preliminaries, we can see *why* (i) and

(ii) are ill-formed. (i) is but a notational variant of

(iv) There is some x such that x is uniquely a king of France and did x go the store.

And (ii) is but a notational variant of

(v) There is some x such that did x go the store?

(v) is quite obviously ill-formed.[79] This fact alone is telling. It is significant that (v)[80] should be so blatantly ill-formed while the sentence that they paraphrase – 'did the king of France go to the store?' – is so obviously well-formed; it is odd – and also discrepant with Russell's theory -- that two supposedly synonymous sentences should deviate so wildly from each other in respect of how well-formed they seem to be.

If a sentence sounds 'funny', as (v) does, that fact is much more likely to alarm linguists than philosophers. Many philosophers will dismiss the 'funniness' of (v) as some relatively unimportant fact about the peculiarities of English grammar, and as not being indicative of any incoherence in Russell's position.

I maintain that (v) sounds strange not (merely) because it is ungrammatical, but because it embodies an incoherence. Given that my appeal to bare linguistic intuitions will not convince many, I must now have recourse to more theoretical considerations to make my case.

For a sentence to have a complete semantics, it is not enough that it encode a proposition: it must also have a *force*. (The sentences 'did John go the store?' and 'John went to the store' are associated with exactly the same

proposition. But those sentences are semantically different; for they have different forces.) A force-less sentence would surely be quite incomplete in respect of meaning and would thus be semantically incomplete. So if (ii) is to have a complete semantics, then it must have a force.

What is the force of (ii)? Let's approach this question indirectly. (ii) is just like (i), except that (i) contains the clause 'x is uniquely a king of France', whereas (ii) does not. With that one qualification, (i) and (ii) are exactly the same. So, in particular, the force borne by the existence-operator (*there is some x such that*) is the same in both (i) and (ii). What force does that operator bear in (i)? Certainly not *interrogative* force. (i) does not mean *is there is some x such that*…? It means *there is some x such that*…So, since the existence-operator plays the same role in (ii) that it does in (i), (ii) means *there is some x such that.*…(not *is there some x such that.…?*) So, in (ii), the existence operator has *assertoric* force.

What follows it, however, has interrogative force. But the thing that has interrogative force is not a proposition but merely a propositional *function* (*did x go the store?*[81]). And the thing that is assigned assertoric force (*there is some x such that*) is not a proposition (or even a propositional *function*) of any kind. So nothing has been asserted: for one cannot assert *there is some x such that.* And nothing has been asked: for one cannot ask *did x go to the store?*

Of course, *if* someone were to come up to me and ask/state (i), I would know well enough what they meant, namely *there is someone who is uniquely a king of France, and I wish to know whether that person went to the store.* But it doesn't follow, and it isn't the case, that this is what is *semantically encoded* in (i).

But you have yet to respond *squarely* to my original point. So I renew it.

One can certainly ask ‘did x go the store’ for some *value* of ‘x’. Mightn’t this be what is going on with (i) and (ii)? A comparison might help. Consider the sentence

(iv) ‘for some x, necessarily x is even.’

This is meaningful (for it is true). Here what is being asserted is not ‘necessarily x is even’ *simpliciter* (that couldn’t be asserted, since it contains a free-variable). What is being asserted is ‘necessarily x is even’ *for some value of x.:* and the result is that (iv) is meaningful. So, by analogy, what falls within the scope of the question-operator in (i) and (ii) is an open sentence; but the sentence as a whole is well-formed. Basically, I don’t see why you can’t quantify through a force-operator.[82]

Force-operators (supposing that such things exist) are not on the same level as modal operators or tense-operators or any other paradigmatic operators. The difference is basically this. Verbiage can fall *outside* the scope of a tense-operator, a modal operator, an epistemic operator, and so on, and *still be meaningful.* But this is *not* true of force-operators.

Every meaningful sentence has a scope-operator and therefore has a *widest* scope operator (one that takes wider scope than all the others): and plainly nothing can fall *outside* the scope of that operator. Everything in a meaningful sentence must be part of an assertion, a question, and so on, and anything that fell *outside* the scope of a sentence’s widest scope operator would *ipso facto* not be a part an assertion or a question…and so wouldn’t have a complete semantics.

Given this we can substantiate our intuition that ‘there is some x such that did x go to the store’ is ill-formed. Let **A** be the assertion-operator and **Q** be

the question-operator. The objector wishes to say that (ii) ‘there is some x such that did x go the store’ – or, using our new symbolism, ‘there is some x such that **Q** x went to the store’ -- is strictly analogous to (iv) ‘there is some such that necessarily x is even’. But there is a difference. In (iv) there is no difficulty giving a semantics to the verbiage lying outside the scope of the necessity operator; for verbiage doesn’t have to fall within the scope of such an operator to be meaningful. But there *is* difficulty giving meaning to the verbiage lying outside the scope of the question-operator in (ii): it must within the scope of *some* operator, lest it not have a complete semantics. The only operator that could conceivably do the trick is the assertion-operator. So we end up with

(v) **A**[there is some x such that] **Q** [x went to the store].

But (v) is patent nonsense. ‘There is some x such that’ cannot *by itself* fall within the scope of a force-operator. This fragment cannot be asserted at all: it cannot be asserted *for* some value of ‘x’ and it cannot be asserted *simpliciter*. (v) is semantic garbage, and the objector’s point misfires.

To close the present argument: (v) makes no sense; but plainly the question whose meaning it is supposed, by Russell’s theory, to give *does* make sense. So (v) doesn’t give the right meaning.

There is an additional reason to regard as meaningless the sentence ‘there is some x, such that x is uniquely a king of France and did x go the store.’ Although this sentence is not itself a conjunction, it is meaningful only if conjunctions of assertions and questions are meaningful.

Consider the open formula ‘x is uniquely a king of France and did x go the store?’. If (iii_F) is meaningful, then it must be possible to replace both occurrences of ‘x’ with a singular term, and thus to have a meaningful

conjunction of an assertion and a question (e.g. ‘Bob is uniquely a neighbor of yours and does Bob drink?’).

Obviously (to use a different example)

(ii) Bob went to the store. Did Bob buy some cheese?

is meaningful. But it isn’t clear whether these two sentences can be fused into anything that is meaningful and deserves to be regarded as a *single* well-formed sentence. Of course we can put an ‘and’ between them. But in so doing, have we really fused them into a *single* well-formed sentence that is a question and an assertion?

First of all ‘Bob went to the store and did Bob buy some cheese’ *sounds* strange.[83] Combination assertion-commands also sound bizarre; no one ever hears: ‘I am hungry and get me some food.’ What one hears is: ‘I am hungry, so you should get me some food’; or ‘I am hungry. [Full stop] Get me some food.’ This shows that we have some kind of pre-theoretical aversion to question-assertions (and command-assertions…); that we don’t intuitively see them as well-formed.

This pre-theoretical intuition can be given some substance. *Part* of the meaning of a sentence (or sentence-token) is a proposition. So if ‘Bob went to the store and did Bob buy some cheese’ is a single sentence, then it has some proposition as part of its meaning. (Presumably that proposition is ‘Bob went to the store and Bob bought some cheese.’[84]) However, associating a proposition with a sentence does not *by itself* give it a complete semantics: the sentence must also be associated with a *force*. ‘Bob bought some cheese’’ and ‘did Bob buy some cheese?’ are associated with the same proposition. But they are semantically very different; for they have different

forces: one is a question and one is an assertion.[85] It is inconceivable that a well-formed sentence should be 'force-less'. If a sentence were 'force-less', then a literal usage of it would never be an assertion or a question of a command or an optative: in effect, such a usage, would not, semantically speaking, be *anything.* And if a sentence is such that a usage of it is never, literally, a question or an assertion…then it is hard to see how that sentence could be viewed as well-formed.

In light of all this, let us consider the sentence 'Bob went to the store and did Bob buy some cheese?' Obviously there is no difficulty associating a force with 'Bob went to the store' and with 'did Bob buy some cheese?' What is in question is whether these two sentences can be meaningfully *conjoined*; whether we can put an 'and' between them with the result that we have a *single* well-formed sentence. Given what we've just noted about force, this doesn't seem possible. First of all, if 'Bob went to the store and did Bob buy some cheese?' is well-formed, then the 'and' must be part of something which has a force. The 'and' must be a part of something a literal usage of which could be the making of an assertion or the asking of a question or the issuing of a command…

But there is no way to make the 'and' a part of anything that has a force. For 'Bob went to the store *and*' cannot, as a whole, be assigned a force. And '*and* Bob went to the store' cannot, as a whole, be assigned a force.[86] (Obviously 'Bob went to the store' be assigned a force; and so can 'Bob went to the store and Bob bought some cheese'; but not '*and* Bob went to the store' or 'Bob went to the store *and*'.) The sentence as a whole cannot be assigned a force – it cannot be said to be an assertion or a command or a question or an optative.[87] So the 'and' cannot be made part of something that has a force – cannot be made a part of something that has a complete

semantics. Thus the ‘and’ is uninterpretable. So the sentence in question (‘Bob went to the store and did Bob buy some cheese?’) contains a constituent that is not given a complete semantics; therefore that sentence itself does not have a complete semantics: it is not well-formed.

I should respond to another objection to this line of reasoning:

> You’ve argued that no-well formed sentence can simultaneously be an assertion and a question, or a command and an assertion. But this is probably false. Consider the sentence: ‘if Mary comes home, should we go to the store?’ or ‘if Phil doesn’t pay the money he owes, repossess his car.’ The ‘if’ part of these sentences is indicative; but the second part is either interrogative or imperative.

This objection is very superficial. Consider the sentence:

(i) If Mary comes home, should we go to the store?

This really means:

(ii) Is it the case that [if Mary comes home we should go to the store]?

In other words, to ask (i) is to ask whether some conditional proposition is true. It is not to *conditionally* ask some *un*conditional proposition. If I say ‘if Mary comes home, then should we go the store?’ I am unconditionally succeeding in asking some question. It is not the case that *if* Mary *fails* to come home, I have *failed* to ask a question. My asking a question is not conditional on Mary’s come home. But suppose Mary doesn’t come home: it is not as though I will in that case have failed to ask a question. It is not

the case that *if* Mary comes home, *then* I am asking whether we should go to the store. (So we should not think of (i) as having the meaning: *Mary comes home → Is the following the case: we should go to the store)* Rather, I am asking if a certain conditional proposition is true; I am asking whether the following proposition is true: *if Mary comes home, we should go to the store.* So we might think of (i) as consisting of the proposition *if Mary comes home, we should go to the store* prefaced by a question-operator (some expressions like 'is the following the case…?')

An exactly similar point holds of apparent command-assertions. Consider the sentence,

(iii) If Bob comes home, then tell him to call me.

Suppose I utter (iii). It is not as though, if Bob *fails* to come home, then I have failed to issue any kind of a command. In uttering (iii) I have succeeded in issuing a command *regardless* of whether Bob comes home. So we must not think of (iii) as being of the form [*Bob comes home]→[make it the case that: Bob calls me]*. The 'make it the case' operator has to be given 'wide-scope'. Otherwise we have the absurd result that, unless Bob homes home, I haven't commanded anything.

So we must think of (iii) as, in effect, meaning:

(iv) Make it the case that [if Bob comes home, then you tell him to call me].

So in uttering (iii), one is not 'conditionally' issuing a command; one is not issuing a command *only* on the condition that Bob makes it home. In uttering (iii) one is *unconditionally* issuing a command. But the proposition commanded is itself conditional in form: that proposition is *if Bob comes*

home, you will tell him to call me. So (iii) can be thought of as a conditional proposition -- *if Bob comes home, you will tell him to call me* – preceded by a command-forming operator: 'make it the case that the following proposition is true'. [88]

What we said about 'and' generalizes to all connectives; to (for example) 'if…then…' Suppose that

(iii) If Bob comes home, then tell him to call me

were a combination indicative-imperative sentence. In that case, by hypothesis, it wouldn't be possible to assign a force to (iii) as a whole. So to give (iii) a force, one would have to assign forces to its sentential components: to 'if Bob comes come' and to 'then tell him to call me'. But 'if Bob comes home' is not *by itself* well-formed and thus cannot be assigned a force; and 'then tell him to give me a call' is not well-formed and cannot be assigned a force. Now given a sentence of 'if-then' form, linguistic usage permits us to drop the 'then' from the consequent; we can say 'if Bob comes home, he will be happy'. But we must think of the 'then' as implicit in such cases. In any case, taken by itself, 'if Bob comes home' is ill-formed and thus cannot, by itself, be assigned a force. But of course, the 'if' must be a part of something that has a force. The only way to make the 'if' a part of something that has a force is to give the sentence *as a whole* a force (that of an imperative). So the sentence as a whole must be seen as a command to the effect that some conditional be made the case, and *not* as something that *will* be a command *provided* that some condition is met. (It makes no sense to suppose that whether I am issuing a command *now* depends on whether Bob comes home three days from now.)

In general, a connective can be made a part of something that has a force

only if *both* sentences (or sentence-formulae) that are joined by that connective are assigned *the same* force. If this condition is not met, then the connective is left 'out in the cold': it is not a part of anything that has a force and is thus not given a complete semantics.

IV. At first Russell's theory *seems* not to do a very good job with imperatives. This has been noted many times.[89] In this section, I'd like to show that Russell's theory really *doesn't* do a good job with imperatives. What I say here will exactly parallel what I said in the last two sections about D-questions.

Consider the following imperative:

(i) Incarcerate the man who wants to kill me.

There are two possible Russellian paraphrases of this sentence, depending on whether we given the definite description wide or narrow scope. The narrow scope reading is:

(ii) Make it the case that exactly one person wants to kill me, and incarcerate that person.

More formally:

(ii$_F$) Make it the case that there is some x such that for any y, y=x iff y wants to kill me, and incarcerate x.

The wide-scope paraphrase is this:

(iii) There is exactly one person who wants to kill me: incarcerate that person.

Formally, it is:

(iii_F) There is some x such that given any y, y=x iff y wants to kill me, and incarcerate x.

Two issues arise. First, it is clear that if Russell's theory is right, then (assuming that the wide-scope paraphrases of D-imperatives are well-formed), when you utter (i), you are either *ordering* that exactly one person want to murder you or you are *asking* whether exactly one person wants to murder you. But intuitively this does not seem to be the case. In a moment I will try to show that this intuitive misgiving is in fact correct

Second, it must be asked whether (iii_F) is meaningful at all. Of course (iii) is meaningful. But (iii) is only a rough paraphrase of (i); it gives the macro-structure, not the micro-structure, of the Russellian reading of (i). It is (iii_F) that gives the micro-structure of (i), and is thus the real Russellian paraphrase of that sentence. Now (iii_F) is peculiar: at the level of literal meaning, it is both an order and an assertion. I will try to show that such combination order-assertions are semantic monstrosities – they are nonsense.

Consider the sentence:

(i) 'Incarcerate Smith's murderer.'

Russell's theory recognizes two paraphrases of this sentence: wide-scope and narrow-scope. For now, let us assume that the wide-scope paraphrase of this sentence *is* well-formed, and let us assign it the proposition that, given this assumption, it is most plausibly seen as bearing (namely *exactly one*

person murdered Smith: incarcerate that person.) Given this assumption, (i) is ambiguous, according to Russell's theory, between:

(ii) *Make it the case that exactly one person x murdered Smith and that you incarcerate x.*

and

(iii) *Exactly one person murdered Smith: incarcerate that person. [Exactly one person x murdered Smith: incarcerate x.]*

But neither (ii) nor (iii) is a meaning borne by (i). Imagine the following. Smith has been murdered. It isn't generally known how many people murdered Smith. A is the detective in charge of bringing Smith's murderer(s) to justice. B knows how many people murdered Smith and, being a friend of Smith, B wants Smith's murderer brought to justice. A knows both of these facts about B, but A does not himself know how many people murdered Smith. Given these facts, consider the following dialogue:

A: Exactly how many people murdered Smith?

B: *__Incarcerate Smith's murderer.__

First of all, B's response is obviously odd. But if Russell's theory were correct, then given PCD, B's response wouldn't be odd. According to Russell's theory, the boldfaced part has as one of its meanings *exactly one person murdered Smith: incarcerate that person.* In the context in question, *if* the boldfaced part *could* be read as having that meaning, it *would* be so read

(this is what PCD tells us). And if B's response were read that way, then it wouldn't be odd. But it is odd. So it doesn't have as one of its meanings: *exactly one person murdered Smith: incarcerate that person.* This is inconsistent with Russell's theory. So Russell's theory is wrong.

According to Russell's theory, 'incarcerate Smith's murderer' has as one of its meanings: *make it the case that exactly one person x murdered Smith and that you incarcerate x.* By an argument similar to that just given, it can be shown that Russell's theory is wrong to assign this proposition to that sentence. Suppose A is a corrupt police-chief and B is his henchman.

A: I really hate Smith. So I want someone to kill him. But I want no more than *one* person to kill Smith; for I cannot spare more than one hit man at the present moment. Also, after Smith is murdered, I want you to incarcerate the person who killed him: that way we will get credit for doing good police work.

B: I have a feeble mind. So could you briefly recap all those instructions, to impress it upon my memory?

A: Very simply: ***Incarcerate Smith's murderer.**

A's response is obviously very odd. If Russell's theory were correct, then given PCD, A's response *wouldn't* be odd. If the boldfaced part did have as one of its meanings *make it the case that exactly one person murders Smith and incarcerate that person*, then competent speakers would read it that way. For reading it that way would completely normalize A's response. But no one *does* read A's response that way. So A's response isn't capable of such a

reading. This is contrary to what Russell's theory says. So that theory is wrong.

Now I'd like to show that Russell's theory assigns nonsense to perfectly meaningful imperatives. Once again, consider the sentence

(i) Incarcerate the person who wants to kill me.

The *true* (fine-grained) wide-scope Russellian paraphrase is

(iii$_F$) There is some person x such that given any man y, y=x iff y wants to kill me, and incarcerate x.

So far we have been operating on the assumption that (iii$_F$) is well-formed. But (iii$_F$) is *ill*-formed. What we said about Russellian wide-scope paraphrases of D-questions applies squarely to wide-scope Russellian paraphrases of D-imperatives.

'There is some person x such that x uniquely wants to kill me, and incarcerate x' is not itself a conjunction. But if it is meaningful, then conjunctions consisting of an imperative and an indicative must be meaningful. (This is because 'x uniquely wants to kill me and incarcerate x' is formed by taking some singular sentence – like 'Bob uniquely wants to kill me and incarcerate Bob' -- and replacing one of the nouns in it with a variable.) But such conjunctions are demonstrably meaningless: the conjunction 'Bob uniquely wants to kill me and incarcerate Bob' is meaningless.

It is not hard to see why. If that sentence were meaningful, it would be possible to assign it, not only a proposition, but also a force. The sentence as a whole cannot be assigned a force; for, as a whole, it isn't an assertion or a

command or a question or an optative. Moreover ‘Bob wants to kill me and’ cannot be assigned a force, and neither can ‘and incarcerate Bob’. So the ‘and’ cannot be made a part of anything that has a force or, therefore, of anything that can be given a complete semantics.

Let us sum up. Either the Russellian wide-scope paraphrase of a D-imperative is well-formed or not well-formed. We’ve seen that, if that paraphrase is well-formed, then Russell’s theory is wrong. And we’ve seen that, if it is not well-formed, then Russell’s theory is wrong. So Russell’s theory is wrong.

I would like to end this paper by considering a couple of objections that might be made to its logical structure. Earlier I gave the following argument. “The king of France” has the same semantics in both

(*) “the king of France went to the store”

and

(**) “did the king of France go to the store?”

Therefore, if Russell’s theory gives the wrong semantics for “the king of France” in (**), then it gives the wrong semantics for “the king of France” in (*), and is thus false *tout court.*

Here is an objection to that argument.

Non-declarative sentences can always be translated into declarative sentences. Thus definite descriptions occurring in non-declarative sentences are, in effect, equivalent to definite descriptions occurring in declarative sentences. “Is the king of France bald” is equivalent to “I wish to know

whether the king of France is bald." And "make the king of France go to the store" is equivalent to "I order you to make the king of France go to the store." If this is right, then for Russell's theory to be correct, all that is necessary is that definite descriptions in *declarative* sentences be amenable to Russell's analysis. And if *this* point, in turn, is right, then the argument of yours that you just outlined loses its force. [90]

The point just considered is based on a fallacy. It is true that, in certain contexts, "did the king of France go the store?" is *pragmatically* equivalent to "I wish to know whether the king of France went to the store." In general, it is true that, at least in some contexts, a question (or an imperative) is *pragmatically* interchangeable with an assertion. But no question is *synonymous* with an assertion. If I say "I wish to know whether the king of France went to the store", I am *saying* something, at the level of literal meaning, about my own mind; I am making an autobiographical statement. But if I say "did the king of France go to the store?", I am clearly not *saying* anything about my own mental state. It is true, of course, that by asking that question, I am *showing* you something about my mental state – I am showing you that I wish to know whether the king of France went to the store. But showing is not saying. When I speak English to you, I am *showing* that I speak English, but I am not *saying* it (except in the special case where I actually *say* "I can speak English"). So there is definitely no synonymy – no *semantic* equivalence - between "I wish to know whether the king of France went to the store" and "did the king of France go the store?", even though those sentences have a certain pragmatic equivalence. [91]

There is one other objection I should respond to. [92] I have said that, if Russell's theory gives the wrong semantics for definite descriptions occurring

in non-declarative sentences, then it gives the wrong semantics *tout court* for definite descriptions. But

"Why should the supposed failure of Russell's theory to give the meaning of definite descriptions in imperatives and questions imply that there is a problem for definite descriptions in assertions? Couldn't the reverse argument that if the definite descriptions in assertions are satisfactorily explained by the theory of descriptions, then the same should hold of those contained in imperatives and questions? Why reason from the failure of the theory to treat d-questions and not the other way around, especially given that many theories have problems with questions and imperatives?" [93]

It would be deeply arbitrary to suggest that "the king of France" has different meanings in "is the king of France bald?", "the king of France is bald", "make it be the case that the king of France is bald", and so on. Surely we want "the king of France" to have a semantics that is uniform across all these contexts. Now if it can be shown that in *one* of these contexts "the king of France" doesn't conform to Russell's theory, then given the point just made, it follows that Russell's theory doesn't give the semantics of this expression as it occurs in *any* of these contexts.

The following might clarify things. It would be deeply arbitrary to suggest that physical objects moving at moderate speeds obey one set of laws while those moving at speeds close to that of light obey a totally different set of laws. Now we know that Newtonian physics gives more or less the right results for bodies moving at moderate velocities. We also know that it gives the wrong results for bodies moving at speeds close to, or coincident with, that of light. The right response to this situation is *not*: "well, since

Newtonian physics gives the right results for bodies whose speed is negligible compared to that of light, it must give the right results *tout court*." No – the right response is: "since Newtonian physics gives the wrong results for bodies moving at speeds approaching that of light, and since bodies moving at high speeds obey the same laws as those moving at slow speeds, we must conclude that Newtonian gives the *wrong* laws (even though it is *approximately* correct within certain contexts)." Similarly, given that definite descriptions have a semantics that is uniform across different contexts, and given that Russell's theory gives the wrong analysis in interrogative contexts, the appropriate conclusion is: Russell's theory gives the wrong analysis *tout court.*

Bibliography

Meaning and Truth First Edition edited by Jay L. Garfield and Murray Kitely Paragon House 1991

The Philosophy of Language Second Edition edited by A.P. Martinich Oxford University Press 1990

Bach, Kent

[1] *Thought and Reference* Clarendon Press Oxford 1987

Blackburn, Simon

[1] *Spreading the Word* Clarendon Press Oxford 1984

Dummett, Michael

[1] *Frege: Philosophy of Language* Second Edition Harvard University Press 1981

Evans, Gareth

[1] *The Varieties of Reference* John McDowell ed. Clarendon Press Oxford 1982

Grice, Paul

[1] 'Presupposition and Implicature' in *Studies in the Ways of Words*
[2] 'Logic and Conversation' in *Studies in the Ways of Words* Harvard 1989

Kaplan, David

[1] 'Demonstratives' in *Themes from Kaplan*
Oxford University Press 1990

Kripke, Saul

[1] 'Speakers Reference and Semantic Reference'
in *Meaning and Truth*

Lambert, Karel

[1] ‘A Theory of Definite Descriptions’ in Ostertag [1].

MacCawley, James D.

[1] *Everything that Linguists Have Always Wanted to Know about Logic* [*But were afraid to ask]* Second Edition University of Chicago Press 1993

Neale, Stephen

[1] *Descriptions* Massachusetts Institute of Technology 1990

Ostertag, Gary ed.

[1] *Definite Descriptions*: *A Reader* Bradford Books M.I.T. 1998

Russell, Bertrand

[1] ‘On Denoting’ in *The Philosophy of Language*

Strawson, Peter

[1] ‘On Referring’ in *The Philosophy of Language*

Counterfactuals: The Epistemic Analysis

Abstract: Ordinarily counterfactuals are seen as making statements about

states of affairs, albeit ones that hold in merely possible or alternative worlds. Thus analyzed, nearly all counterfactuals turn out to be incoherent. Any counterfactual, thus analyzed, requires that there be a metaphysically (not just epistemically) possible world w where the laws are the same as here, and where *almost* all of the facts are the same as here. (The factual differences relate to the antecedent and consequent of the counter-factual.) But, as I show, this requirement typically involves the positing of worlds whose necessary non-existence can be shown by fairly elementary deductions. Further, the possible-worlds analysis of counterfactuals is guilty of covert circularity. For, thus analyzed, counterfactuals can only be understood in terms of laws of nature (the laws that apply here are assumed in the hypothetical world – except in the very special case where the counterfactual is also a counter-nomic). But there is widespread agreement that the concept of a law cannot itself be defined except in terms of the notion of a counterfactual (a law is *inter alia* something that supports counterfactuals). I give a purely epistemic analysis of counterfactuals, arguing that they are crypto-probability propositions. I also argue that the relevant kind of probability can be defined wholly in terms of what *has* happened (not what *would* happen and not even what must happen in a nomic sense); so that my analysis isn't guilty of any kind of circularity.

I. Statements like the following are examples of *counterfactuals*:

(i) If Kennedy had not been assassinated, the U.S. would have pulled out of Viet Nam before 1965.

(ii) If Jones had not fired the gun, Smith would still be alive today.

A counterfactual says what *would* have been the case *if* some condition C obtained, when in fact C *did not* obtain.

The consensus is that no acceptable analysis of counterfactuals has been proposed. I agree with the consensus. The reason there is no satisfactory analysis is that the existing analyses regard counterfactuals as saying something about causal relations among *states of affairs* – albeit ones that take place in non-actual or actual but alternative worlds. The right approach, I argue, is to see counterfactuals as expressing relations of probabilification holding among *statements*. Counterfactuals don't really say anything about *events* or anything concrete – or even anything that *would* be concrete if it existed. Counterfactuals merely register confirmation relations among propositions. That is what I will try to show in this paper.

I must make one thing clear. I am taking it for granted that counterfactuals are, or can be, true. I am not arguing that they are necessarily incoherent or otherwise false. So *given* that a factualist analysis makes them incoherent, a non-factualist analysis – an epistemic one – is needed.

Of course, it has been argued (by Quine, for example[94]) that counterfactuals *are* meaningless or, at least, systematically false. But given how endemic they are to our thought, it would be better to find a way to make them non-meaningless than to try to do without them.[95] (I am open to the idea that they are all meaningless or false, but I will try to show how we can avoid coming to that undesirable conclusion.)

II. First of all, what is meant by an 'analysis' of counterfactuals? An analysis of a counterfactual says *in virtue of what* some counterfactual holds. Suppose that (ii) is true. Surely it is not a *brute* fact that, if Jones had not fired

the gun, Smith would still be alive. *Prima facie* (ii) seems to make a statement about something non-actual (one non-actual state of affairs is followed by another). But surely the truth of (ii) derives from what is *actual*. In other words, if (ii) is true, that is wholly in virtue of some fact about what *is* the case – facts about the laws governing our universe and also, perhaps, about Smith's constitution (he is not made of steel, but of flesh). An analysis of counterfactuals identifies the *actual* facts in virtue of which some counterfactual holds.

This no doubt demands further clarification. Let w and w* be two parallel universes that are exactly the same in every respect (except they are numerically distinct): the same states of affairs obtain, and the same laws link those states of affairs. Suppose that, in w, Bob says 'if Smith hadn't smoked so much, he wouldn't have gotten emphysema.' In w*, Bob* (Bob's counter-part in w*) says the exact same thing. Surely Bob's statement is correct if and only if Bob*'s statement is correct. It isn't possible that what is counterfactually true with respect to w should diverge from what is counterfactually true with respect to w*. And the reason, surely, is that once the *actual* facts – the actual states of affairs and the actual laws that link them – have been laid down, the *counterfactual* facts (or truths) are thereby fixed. The counterfactual 'supervenes' on the factual. So an 'analysis' of counterfactuals reduces the counterfactual facts to the actual facts.

David Lewis has given the most perspicuous account of the conceit that underlies ordinary thinking about counterfactuals, and that underlies virtually all philosophical thinking about it:

'If kangaroos had no tails, they could topple over' seems to me to mean something like this: in any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as having no

tails permits it to, the kangaroos topple over.[96]

Notice that Lewis construes counterfactuals as making statements about *states of affairs* – albeit ones that hold in worlds other than this one. What makes (ii) true is (inter alia) that, in some other world, Jones doesn't fire the gun and Smith still remains alive: what makes (ii) true is that certain sequences of affairs (Jones' trigger finger being in certain positions and not in others) are succeeded by other states of affairs (Smith's heart continuing to beat, his brain continuing to function, and so on). So (ii) makes a statement about states of affairs or *facts*, albeit ones that hold in other worlds. If Lewis is right, what (ii) does *not* do is (merely) make a statement about logical relations among *propositions*: (ii) is about facts slapping against facts, bodies hitting (or not hitting) other bodies; it isn't about a shadow world of logical interrelations holding among platonic entities. Let us describe as *factualist* any analysis that is, in this respect, like Lewis'. The factualist says that counterfactuals are about states of affairs, facts, and objects – the things we encounter in this world – except that they are in merely 'possible' worlds or in alternative worlds. I will advocate a *non*-factualist approach: one that is depressingly anemic compared to Lewis'.

Notice, also, that Lewis has done a good job of reducing the counterfactual to the actual. For his analysis is essentially this. *If kangaroos had no tails, they would topple over* means: in a situation S where kangaroos *don't* have tails, but where things are otherwise just like our situation (so far as that is possible, given that kangaroos don't have tails in S), kangaroos *do* topple over. Lewis has replaced the subjunctives with indicatives. What *would* be the case here is what *is* the case somewhere else. So he has reduced what *would* be the case to what *is* the case. And this is just what an analysis of

counterfactuals is supposed to do. (Of course, Lewis has accomplished this reduction by broadening the actual to include what would *ordinarily* be considered non-actual. And he has been taken to task for this. But his solution does embody a keen awareness of the need for any analysis of counterfactuals to ground the truth of them[97] in what is actual: and to that extent he deserves credit.)

In his factualism[98], Lewis is, it seems to me, very much in agreement with the ordinary person. When the ordinary person says 'if Smith hadn't pulled the trigger, Jones would still be alive', he thinks that his statement is true in virtue of some relation among states of affairs or events – with the qualification that these events didn't happen in the actual world, but in some possible world. The ordinary person definitely does *not* think that he what is he saying is true in virtue of relations of confirmation among shadowy entities like propositions.

But the factualist view is not tenable. That is what I will try to show. Counterfactuals do not say anything about causal (or any other) relations among possible, but non-actual, events. And they don't say anything about relations among events that are in worlds other than this one. Counterfactuals say something *only* about the degree of credibility that certain propositions confer on other propositions.

In any case, *if* counterfactual propositions have any chance of being true, they must be seen this way. Otherwise, counterfactual propositions are systematically logically false --in the same category as *p and not p.*

I don't dispute that Lewis is right about what people are *thinking* when they make counter-factual claims. But in so far as Lewis is right about that, then what people are thinking is totally incoherent. I will produce something that is as close as possible to the ordinary way of thinking *while still being*

coherent.

III. Suppose that, in actual fact, Smith shoots Jones and thereby kills him. Given this, Let us consider the proposition:

(ii) If Jones had not fired the gun, Smith would still be alive today.

This is a perfectly good counterfactual; and, if we fleshed it out a bit more, it is the kind of counterfactual that we might well regard as true. Let us apply the factualist analysis to it.

That factualist analyses (ii) as meaning, or being equivalent to,

(iii) Consider a world – call it w -- with the following property: in w, Jones doesn't pull the trigger; but, with that qualification, w is as much like our world as it could be. In w, Smith is still alive.

The factualist must also say something about the *laws* obtaining in w: he must say that those laws are the same as the laws that govern *our* universe. For if the physical laws of w are *different* from those in our world, then the counterfactual is immediately voided. If those laws are different, then Jones' pulling the trigger might not lead to the bullet leaving the gun, or Smith's flesh might repulse the bullet, or there might not even be a gun – perhaps those other laws wouldn't permit the atoms in question to bind to one another in the right way. So clearly w must be governed by the same laws as our world.

Also, it seems plain that when one makes a counterfactual assertion like (ii), one is implicitly assuming that, in the counterfactual situation, the same laws apply that apply here: this seems to be essential to the very idea of a

counterfactual. It doesn't matter what one's theories on counter-factuals are. When I say

(*) If Kennedy hadn't been shot, LBJ would not have become president in '63,

I am obviously saying or, at least, committing myself to:

(**) if Kennedy hadn't been shot, LBJ would not have become president in '63 -- *provided* of course that some cataclysmic shift in the laws of nature didn't occur in '63.

Obviously if there had been some shift in the laws of nature in '63, then all bets are off: maybe that shift in laws would have made LBJ become president in '63.

A counter-*nomic* varies the laws; a counterfactual varies the facts, but not the laws. The whole point of a counterfactual like *if Clinton had resigned, Gore would have been president* is that *given* the laws operative here, a difference in initial conditions (Clinton's resigning as opposed to not resigning) could only have led to a different outcome (Gore's being president as opposed to not being president).

Given all of this, let $L_1 \ldots L_n$ be the laws that govern our world. As we just saw, $L_1 \ldots L_n$ must also govern w. Now in our world, $C_1 \ldots C_n$ led to Jones' pulling the trigger. Of course, $C_1 \ldots C_n$ led to that event because laws $L_1 \ldots L_n$ govern our world. (A different set of laws needn't lead from those initial conditions to that particular conclusion.) Trivially, laws are either deterministic or they're not. A 'deterministic' law says: given such and such

conditions, thus and such happens *100%* of the time. An 'indeterministic' law says: given such and such conditions, thus and such happens *a certain percentage of the time* (not 100%). Let us first investigate the consequences of supposing that $L_1 \ldots L_n$ are deterministic. Then we will consider the consequences of supposing that at least *some* of the laws in $L_1 \ldots L_n$ are indeterministic. (This exhausts all the possibilities: either *all* of $L_1 \ldots L_n$ are deterministic, or some of them are *not* deterministic, i.e. are indeterministic.)

IV. First case: suppose $L_1 \ldots L_n$ are deterministic. In that case, in w, $C_1 \ldots C_n$ lead to Jones pulling the trigger. Now remember that we are trying to keep things in w *as much like things in our world* as is possible given that, in w, Jones did not pull the trigger. We've just seen that we cannot keep $C_1 \ldots C_n$ the same; at least one of them must go. So, for some i, C_i must *not* be the case in w. But C_i did not just come out of nowhere; some *previous* set of conditions $C^*_1 \ldots C^*_n$ are what led to C_i. So given that C_i is *not* the case in w, and given that the laws leading from $C^*_1 \ldots C^*_n$ to C_i are deterministic, we must suppose that, for some i, C^*_i is not the case. Otherwise we are stuck with C_i, and if we are stuck with C_i, we are stuck with Jones' pulling the trigger. Of course, C^*_i didn't just come out of nowhere; it arose from some prior conditions $C^{**}_1 \ldots C^{**}_n$. So we must assume, for now familiar reasons, that for some i, C^{**}_i is not the case. Cleary this regress goes on and on.

In fact, this regress goes back into the furthest recesses of the past. So it is clear that, if Jones is to *not* pull the trigger in w, then w's past must be *very* different from the past of our world. But if w's past is *that* different, then there is no guarantee at all that w's present would even *remotely* resemble our present. What we *want* is for w's present to be just like ours *except* that Jones doesn't pull the trigger. But what we get is a present which almost certainly

doesn't even have a Jones or a Smith and which, even if it does comprise a Smith and Jones, is so different from our world, that nothing can be said with confidence about it.

One might try to fix things in the following way:

Here is what you say. If $L_1 \ldots L_n$ are deterministic, then we can't get Smith to avoid pulling the trigger without changing a lot of things about w's past; and that if all those changes are made to w's past, then there is no reason to expect that w will bear even a *remote* resemblance to our world, let alone that Smith and Jones will be in some particular predicament together.

But there is no need to make changes to w's remote past to get the result we want (namely, that everything in w is just like it is here, except that, in w, Smith does not pull the trigger). We don't even have to tinker with $C_1 \ldots C_n$. We can leave those alone. All we need is that, in w, there is a brief and isolation *suspension* of $L_1 \ldots L_n$. So $C_1 \ldots C_n$ hold in w (just as they do here). Thus we don't need to make any changes to w's past. All we ask for is that, for a brief moment, $L_1 \ldots L_n$ don't apply (at least not to $C_1 \ldots C_n$ – we can let them apply to other regions of w) and that the consequence of their *non*-application is that Smith does not pull the trigger.[99]

Let us refer to this move as M (for miracle). M doesn't work. The basic problem is that it isn't possible to contain or isolate suspensions of causal laws. Suppose we don't want some event E_1 to occur in our counterfactual scenario; and we keep E_1 out by suspending $L_1 \ldots L_n$. What ends up happening is that, because of that suspension, various *other* events $E_2 \ldots E_n$ that we don't want end up happening. So, in suspending $L_1 \ldots L_n$, we end up

committing ourselves to counterfactuals that are not true or that, in any case, are very different from the one we want to assert. In other words, in suspending $L_1 \ldots L_n$ in w, we are opening the door to all kinds of things we don't want to let in; and our counterfactual ends up being to the effect that (say) Jones skeleton turned to gelatin (in which case, the counterfactual in question would become false) or (say) that the house Jones was in disintegrated (in which case our counterfactual says much more than we want it to).[100] To prevent $E_2 \ldots E_n$ (the unwanted fall-out of suspending $L_1 \ldots L_n$) from cluttering up our counterfactual scenario, we have to keep on adding miracle upon miracle. But, by logic similar to that just described, that makes the situation worse.

All of this becomes apparent when we start to think a little more rigorously about the nature of causation. We say that *the* cause of the crash was Brown's inattention to the road. But the truth is that innumerable conditions had to obtain for Brown to crash. The road and the tires had to be made of a certain material; the tread on the tires had to be worn down by a certain amount; the road had to be wet; the car had to be moving with a certain speed; the mechanisms connecting the steering wheel to the rotations of the tires had to have, or lack, a certain degree of integrity; and on and on.[101] Given virtually any event, the conditions that were causally necessary for that event to occur are innumerable.

The other side of the coin is that, if some set of conditions $C_1 \ldots C_n$ is sufficient to bring about a given event, then it is *also* sufficient to bring about innumerable other events. Nothing that causes anything causes just *one* thing.

Given this, suppose that, in w, $L_1 \ldots L_n$ are suspended, so that (in w) $C_1 \ldots C_n$ don't lead to Jones' pulling the trigger. That suspension will not *just* lead to Jones' not pulling the trigger; it will have innumerable other consequences.

It is overwhelmingly likely that what was necessary to get Jones to pull the trigger was *also* necessary for Jones' head not to explode, for the house he was in not to disintegrate, for the electrical activity in Jones' brain not to cease. Among the necessary conditions of Jones' pulling the trigger – in other words, among $C_1 \ldots C_n$ --- are facts having to do with the integrity of the various alloys composing the gun; facts relating to the architecture of the house Jones was in; facts relating to the durability of the materials composing the floor Jones' was standing on; facts relating to the amount of pressure that Jones' vertebrae could withstand without pulverizing; and so on. If we suspend $L_1 \ldots L_n$ just enough to make sure that Jones does not pull the trigger, then we are suspending them quite enough to vaporize Jones' spinal cord, to melt Smith's brain, to turn the house that they are both occupying into a heap of smoldering rubble. So somebody who advocates M is committed to all manner of consequences that someone affirming (ii) does not want.

(ii) is (or could be) a true counterfactual. But surely

(iii) If Jones hadn't pulled the trigger, then the house Smith and Jones are occupying would turn to dust

is not a true a counterfactual. More accurately, if (iii) *is* a true counterfactual, it is a very different one from (ii) – not one that everyone who held (ii) would necessarily wish to countenance.

There is no way to *contain* suspensions of laws. If you give such suspensions an inch, they take a mile.[102] If a miracle – a suspension of physical (or psychological) laws -- is sufficient for one thing, then it will be sufficient for things that are not even *counterfactual*ly true. For it is not even *counterfactually* true that, if Jones' had not pulled the trigger, then the house

would have disintegrated. But that – or, if not that specifically, then some other bizarre consequence – is just what is incurred by an advocate of M.

A single theme has run through everything we've said so far. Any given event is so deeply embedded in other events, that you cannot consider a world that is even *remotely* like our world but doesn't comprise that event. If you get rid of that event, then the counterfactual situation becomes so unlike the actual situation that you either don't get the result you want (there is no Jones or Smith in the first place, and *a fortiori* Smith doesn't live) or, if you do get the thing you want (Jones' not pulling the trigger), you get a lot of things that you didn't bargain for and aren't even counterfactually true (the house disintegrates, Smith's brain melts).

Because of the 'embedded' character of events, and the consequent 'non-containability' of suspensions of causal laws, there is no way to construe counterfactuals as holding in virtue of some kind of relation among *states of affairs*, whether in some possible world or in some actual, but alternative, world. When we run a counterfactual, we must typically[103] assume that the causal laws governing the hypothetical world are the same as they are here. But if we keep the laws the same, and then try to vary some events, the results ramify out of control. It doesn't help to suspend the laws; an undesired ramification still occurs. There is simply no way to construe counterfactuals as holding in virtue of relations among facts in some alternate or possible world. For the 'possible' world that *would* validate the counterfactual is *not* in fact possible: given the physical laws that are implicated in the counterfactual, and given the way we want the initial conditions to be in the counterfactual, it is not possible – I mean: not *logically* possible – for an alternative world that *would* validate the counterfactual exist. It is not *logically* possible because, given the laws and given the kind of variation in events that we need to construct the 'validating' world, what we end up with

as a matter of logic is something in which either the counterfactual is false or in which, if the counterfactual does come out true, many other counterfactuals come out false.

It may seem odd to say that such a world is not *logically* possible; for nothing is more obvious, it would seem, than to say that there are logically possible worlds in which Jones doesn't shoot Smith. There are such worlds. (I mean: they are logically possible. I leave it open whether there is a concrete world corresponding to every possibility.) But the existence of such worlds is not enough to validate our counterfactual claims. For (ii) to be a true counterfactual, it is not enough for it to be logically possible for Smith to not shoot Jones. Smith must *not* shoot Jones in a world that has two key features: (i) that world is governed by the same laws that govern our world; (ii) the states of affairs that hold in that world are *almost* (but not quite) identical with those which hold in our world. These two features are essential to all counterfactuals (except those that happen also to be counter-nomics: a very small class). If the laws in the hypothetical world w aren't exactly like the laws in our world, then there is *no* legitimate inference from 'Jones doesn't shoot Smith in w' to 'Smith doesn't die in w'; for the laws in w might be such that Smith *must* be shot at least once a day to keep on living.

And if the facts in w aren't very similar to the facts in our world, once again there is no legitimate inference from 'Jones doesn't shoot Smith in w' to 'Smith doesn't die in w'. What we want is a situation that is *very much* like the one that actually obtained but that *doesn't* culminate in quite the same ending. We want Jones and Smith to exist; we even want them to have more or less the relationship that they actually have; we might even want them to be the in same room. For if we the facts in w bear *no* resemblance to the facts here, then we have absolutely no guarantee that Smith might not die (albeit for some other reason than by being shot); we have no guarantee that Smith

won't be eaten by a bear or fall off a cliff. In fact, if we construe the events in w as being anything short of *very much like* the states of affairs here, then the counter-factual is immediately voided.

So if we construe counterfactuals as holding in virtue of *states of affairs* in alternate worlds, we have to walk an extremely fine line between keeping things in that world very similar to our world and *not* keeping them *so* similar that the outcome which *actually* occurred ends up occurring over there. And we must walk this fine line *while* keeping the laws in that alternative world the same as they are here.

But it is not *logically* possible to do all three things at once. As Hempel said, time and time again, *given* a set of deterministic laws $L_1 \ldots L_n$, and *given* a set of conditions $C_1 \ldots C_n$ falling under those laws, it *logically* follows that thus and such will happen. Obviously it isn't logically necessary that those particular laws obtain, or that those particular conditions obtain. But it *is* logically necessary that, if those laws obtain *and* those conditions obtain, there will be a particular outcome.

If counterfactuals are construed as holding in virtue of relations among *states of affairs* (albeit states of affairs in alternative worlds), it *must* be possible to do all three things at once. It is not possible (at least typically); so counterfactuals are not to be construed.

But this must be qualified: we have seen that *if* the laws governing the world are *deterministic*, then counterfactuals are not to be so construed. We haven't considered the case where at least some of those laws are *indeterministic* (probabilistic). Let us do that now.

V. *Second case: suppose some laws are indeterministic.* Once again, let $L_1 \ldots L_n$ be the laws governing our world. And suppose that, for some i, L_i is indeterministic.

First of all, it is clear that our counterfactuals are only as good as our laws are deterministic. If being shot in the chest is fatal 100% of the time, then I can say with complete certainty 'if Jones *had* been shot in the chest, he *would* have died.' The counterfactual is as good as the law in question is deterministic (the law being: x is shot in the chest → x dies). If that law is probabilistic, then the counterfactual is that much less credible. If being shot in the chest leads to death 20% of the time, then the counterfactual 'if x had been shot in the chest, x would have died' doesn't deserve much credence at all.

But, of course, given only that *some* laws are indeterministic, it doesn't follow that *all* counterfactuals are thereby vitiated: only those counterfactuals are vitiated which depend (in a certain way) on the probabilistic laws in question.

Also, even if *all* laws were indeterministic, it wouldn't follow that counterfactuals couldn't legitimately be made. For if the probabilities are high enough – say, 99.99999% - then the laws would, in effect, support counterfactuals as well as deterministic ones. (In fact, there is evidence that *all* laws are indeterministic; the laws that *seem* deterministic are actually probabilistic ones, but the probabilities in question are very close to 100%.)

But it is not hard to show that the existence of indeterministic laws doesn't help the factualist at all. It actually makes things worse for him.

There is one very obvious reason for this. To help put this reason in context, let us momentarily forget that we are talking about *in*deterministic laws and let us, once again, talk about deterministic laws; let us once again suppose that $L_1 \ldots L_n$ (the laws that govern our universe) are *deterministic*. It is, at the very least, an epistemic possibility that this supposition should turn out to be true (Einstein thought it *was* true). Taking 'might' in a purely epistemic sense, future science might reveal that all our laws are

deterministic. Now if this should turn out to be the case, it would obviously *not* mean that all our counterfactuals statements would turn out to be false. Consider the following counterfactual:

(a) If Brown had struck the window with the hammer, it would have broken.

Let us suppose that the window is very thin, that Brown is strong, and so on – in other words, let us suppose that (a) is a true counterfactual. If future physicists discover that (e.g.) quantum mechanics is deterministic, would it mean that (a) turns out to be false? Suppose in the year 2135 it is finally established that all physical laws are deterministic. Will people then say: 'I guess we were wrong to say that *if* President Clinton had resigned, Gore would have become president'? Of course not. Those counterfactuals would *still* be true. Few people prior to 1900 believed in indeterministic laws. (The very idea seemed self-contradictory to many.) But that was (rightly) not seen to threaten the possibility of making true counterfactual statements. Basically, an *analysis* of counterfactuals cannot make the truth of such statements depend on things on which their truth is not *in fact* dependent. It is a datum that (a) is true. It is not a datum (though it may be true) that

(b) there are indeterministic laws.

So the truth of (a) cannot be made to presuppose that of (b).

And, as we'll see in a moment, given someone who is relying on the existence of indeterministic laws to validate his analysis of counterfactuals, that person is committed to extremely specific hypotheses as to *which* situations those laws are operative in. He needs them to *not* to be operative in

certain chosen circumstances. (For if they *are* operative there, then he can't be *sure* that, say, the window would have broken if struck: the indeterministic character of the laws involved might allow the window to stay intact.) But he *does* need them to operative in certain circumstances *not* directly associated with that window-striking. (Otherwise, he can't guarantee that there will even be a window or a hammer or somebody available to strike the former with the latter; the indeterministic character of the laws involved might not even suffice for the preconditions for his counterfactual.) So the factualist-indeterminist (i.e. somebody who is a factualist but predicates the truth of counterfactuals on the existence of indeterministic laws) ends up being committed to quite specific beliefs – beliefs which may well be false and which probably are false – as to where and when indeterministic laws are operative. So the factualist-indeterminist makes the true of (a) turn on the outcome of scientific investigations that may not, and probably will not, go his way.

But even if we lave that aside, the factualist-indeterminist is in trouble if he takes some of $L_1 \ldots L_n$ to be indeterministic. A factualist who is running (a) is going to want the laws that mediate between Brown's striking the window and the window's shattering to be *deterministic* or, if not deterministic, then very nearly deterministic; in other words, those laws must pair off some very high percentage of such window-strikings with window-breakings. For, obviously, if those laws assign window-breakings to only 42% of such window-strikings, then the counterfactual won't be true or, in any case, won't have much credibility.

Let $L^*_1 \ldots L^*_n$ be the laws that that mediate between Brown's (hypothetical) window-striking and the subsequent (hypothetical) window-breaking. Of course, $L^*_1 \ldots L^*_n$ are a subset, either proper or improper, of $L_1 \ldots L_n$.

We just saw that the person running the counterfactual needs $L^*_1 \ldots L^*_n$ to be deterministic, or nearly so. But right there he gets into serious trouble on a number of levels.

(Henceforth let us refer to the person running the counterfactual as our 'hypotheticalist'; I don't want to call him our 'counterfactualist', since that would obviously be confusing.)

Why does our hypotheticalist get into trouble there? Trivially, $L^*_1 \ldots L^*_n$ are either a proper subset of $L_1 \ldots L_n$ or they are an improper subset of $L_1 \ldots L_n$. (The third 'possibility' – that $L^*_1 \ldots L^*_n$ is not a subset of $L_1 \ldots L_n$ -- is non-existent since, by assumption, $L_1 \ldots L_n$ are the totality of laws governing the world and $L^*_1 \ldots L^*_n$ are the subset of those laws implicated in some particular circumstance, namely Brown's striking the window.) Let us consider each case. If they are an *im*proper subset (in other words, if $L^*_1 \ldots L^*_n$ are *identical* with $L_1 \ldots L_n$), then it follows that $L_1 \ldots L_n$ are deterministic. But we've already dealt with this case: we've seen that, if $L_1 \ldots L_n$ are deterministic, then the factualist story is *logically* incoherent.

So we must assume that $L^*_1 \ldots L^*_n$ are a *proper* subset of $L_1 \ldots L_n$. Let $L'_1 \ldots L'_n$ the laws that are in $L_1 \ldots L_n$ but *not* in $L^*_1 \ldots L^*_n$. Now if all of $L'_1 \ldots L'_n$ are deterministic, then once again, we get the result that all of $L_1 \ldots L_n$ are deterministic which the factualist cannot have. So we must assume that some of $L'_1 \ldots L'_n$ are indeterministic. Further, we must assume the probabilities in question are significantly less than 100%. For if we say that if these probabilities are close enough to 100%, then we run into the same problems we ran into when supposing that $L_1 \ldots L_n$ are all deterministic. The problem is, basically, this: given how embedded any event is in other events, if $L_1 \ldots L_n$ are close to deterministic, we *logically* can't assume a change in a given event with that change ramifying in ways that will inevitably falsify the

very counterfactual we are running. So $L'_1 \ldots L'_n$ need to be non-negligibly *in*deterministic: they need to dole out nice *low* percentages like 76% and 32%. In other words, $L'_1 \ldots L'_n$ cannot be of the form *n% of F's are G's*, where n is 99.999. They need to be of the form *n% of F's are G's*, where n is some number like 45 or 81.

So let us concede this to our hypotheticalist; let us concede that $L'_1 \ldots L'_n$ are of the form *n% of F's are G's*, where n is some number significantly less than 100. And let us consider the consequences of this concession.

As we've seen, our hypotheticalist needs the laws mediating between Brown's hypothetical window-striking and the subsequent window-breaking to be deterministic, or nearly so. Let w be the world where these hypothetical events happen. The problem is this: in real life Brown didn't strike the window; he did something else. He (let us say) threw the hammer in the trash-can. Our hypotheticalist needs to make sure the laws in w are the same here. As we've seen, he cannot get this result if the laws in w (i.e. the laws here) are deterministic. So they must be indeterministic. The idea is that the 'wiggle-room' allowed by these laws will allow for events that *in fact* led up to Brown's throwing the hammer away to lead up (in w) to his striking the window. Let $C_1 \ldots C_n$ be the conditions that in fact led up to Brown's throwing the hammer away. With regard to the laws that *in fact* mediated between $C_1 \ldots C_n$ and Brown's throwing the hammer away, even if those laws are *in*deterministic, it by no means follows that they are consistent with Brown's striking the window. In fact, even if those laws are *highly* indeterministic, it is still possible that, given $C_1 \ldots C_n$, the probability they assign to Brown's striking the window is zero. Indeterministic laws are not blank-checks; an indeterministic law is not one of the form *given an F, that F could be succeeded by* anything *at all.* A deterministic law leaves *no* options; there is only *one* way things could go. An indeterministic law leaves open

some options, but it doesn't leave open *any* option you please. The options are severely circumscribed. It is (we are told) undetermined when exactly a given electron will jump to another orbit; the best we can do is assign a probability to the electron's jumping orbit at t. So the law in question is probabilistic; there is *no* law (we are told) requiring the electron to jump at t_1 as opposed to t_2. But this doesn't mean that *everything* about the electron's behavior is undetermined. It *is* determined that the electron is not going to jump to an atom that is a mile away; it *is* determined that the electron will not turn into a volleyball. The electron is 'free', so to speak, only within the narrowest of margins. Even believers in indeterministic laws grant that the macro-world deviates from determinism, at most, by a vanishingly small amount. The reason the macro-world is deterministic, or vanishingly close to deterministic, even though the micro-world is (allegedly) indeterministic, is that the latter indeterminacies are circumscribed: some specific aspect of some state of affairs is left open; not *everything* is left open. And when something in the micro-world is left open, it isn't left *totally* open; it is left only in some quite specific respect. The result is that aggregates consisting of things which are not subject to deterministic laws themselves still behave deterministically; for the indeterminacies stay within certain bounds.

The way Company F behaves is highly determined by the Sergeant's orders, even though his orders don't determine whether or not Private Smith's eyelid twitches at some particular moment. The Sergeant's orders ensure that these the indeterminacies stay within certain bounds. So even though a *certain* indeterminacy (relative to the Sergeant's orders) prevails, that does not translate into indeterminacy on the part of the behavior of the F-company taken as an aggregate. Our hypotheticalist needs the indeterminacies in $L'_1 \ldots L'_n$ to be uncircumscribed. But they are not. So our hypotheticalist doesn't get what he wants; he doesn't get Brown to strike the

window. Of course, he *could* try to get what he wants by assuming that the laws in w are different from the laws here. But that, as we've seen, would nullify the counterfactual. For the essence of a counter*factual* – though not a counter-nomic – is that the laws in the hypothetical scenario are the same as in the home universe.

To sum up, there is no way to make counterfactuals say anything about relations between states of affairs or events: not even about events in alternative or possible worlds. The essence of a counterfactual is that the laws that apply in the hypothetical situation are the same as the laws that apply here. If we construe a counterfactual as making a statement about relations between events (or states of affairs), then we have to keep those laws the same while varying some crucial initial condition. But that is *logically* impossible: the laws being what they are, it is *logically* impossible to vary *just* the initial condition in question without varying various others which will almost inevitably make the counterfactual turn out false. The impossibility is logical, and not merely nomic (or even metaphysical), because it is a truth of *logic* that given a law L to the effect that all F's are G's (or 80% of F's are G's), it follows *logically* that a given F is a G (or has an 80% chance of being a G), even though isn't logically necessary that all F's are G's are that that particular thing is an F.

I will end this section by addressing a possible objection:

The argument you just gave seems to involve the premise that the macroscopic world is "vanishingly close" to being deterministic. Admittedly, there is a fair amount of empirical support for this view, and it is a standard one. But it might turn out to be wrong. In that case, it seems, your argument would fail. So your argument presupposes the truth of a dubious empirical doctrine.

For the sake of argument, let's suppose I *am* wrong; let's suppose that the that macro-scopic world is not vanishingly close to deterministic and is, in fact, non-negligibly indeterministic. In that case, counter-factual claims concerning the macro-world will automatically be nullified. Suppose the exact same initial conditions lead to (say) the car starting only 73% of the time. Under these circumstances, if somebody

(*) *if I'd turned the key, the car would have started*

the right response is: 'maybe, maybe not'. So (*) is surely not *true.* Perhaps it is 73% true.[104] More likely, there is no fact as to whether it true. But it is definitely not 100% true. And that is, by most lights, the *only* kind of truth.

Counter-factuals are only as good as the world is deterministic. So even if I am wrong about the macro-world, that doesn't help the factualist; it actually makes things worse for him.

VI. To complete our argument against the factualist, we must consider one other move he might make. To put this move into context, it will help if we briefly sum up what we've said.

The factualist needs there to be worlds that are *almost* exactly like our worlds. For every proposition P that is not true of our world, the factualist needs there to be a world w that is just like our world *except* that P is true in w.

But either this demand cannot be met *or*, if it can be met, the conditions that enable it to be met destroy the factualist's position.

Either our world is deterministic or it is not. When we run a counter-factual, we are considering a world w where the *laws* are the

same as they are in our world.

Two deterministic worlds cannot differ from each other with respect to a single state of affairs unless they differ with respect to a series of states of affairs extending back into the furthest recesses of time. This is a consequence of the fact that, in a deterministic world, no event comes out of nowhere. So any difference between two deterministic worlds must traced to some prior difference D1; and D1 must in its turn be traced to some prior difference D2; and so on. Thus a single difference must be grounded in an indefinitely long history of differences.

So if Jones shot Smith in w, and Jones didn't shoot Smith in this world, then this world and w will differ in unfathomable ways. We thus have absolutely no idea what the consequences of Jones' shooting Smith in w might be. So w doesn't validate the right counter-factuals and it validates the wrong counter-factuals.

So if our world is deterministic, then the factualist approach is a failure.

Granted, two *in*deterministic worlds may indeed differ from each other in some one respect. But counter-factuals are sustainable only in so far as the world is deterministic. So to the extent that couinter-factuals *can* be truly affirmed, the world is deterministic: and we run into the problem just described. We avoid that problem problem only in so far as the world is *in*deteminstiic. But in so far as it is indeterministic, it doesn't sustain counter-factuals at all, no matter how they are construed.

So the fatualist analysis fails if the world is deterministic, and it fails if the world is not detrerminstic. So it fails.

But there is one move that the factualist might still make:

Suppose that in actuality, Jones does not shoot Smith. And suppose that Jones is psychologically normal and thus *would* feel remorse if he did something as heinous as shooting somebody.

Now consider the counter-factual:

(*) If Jones had shot Smith, Jones would now be feeling profound remorse.

Let w be some world that is just like ours except for this one thing. In w, at time t, Brown shoots Smith. In the actual world, at time t, Brown does not shoot Smith.

I grant that the factualist runs into the problem you desribed *if* he insists on tracing every difference between w and our world to some pre-existing difference between them. For then, obviously, he will have to suppose that w's entire history deviates from ours.

But there is no reason why the factualist must do that. The factualist can say this. Suppose that *up until t* w and the actual world are exactly the same. But at t, the two worlds diverge. In w, Jones shoots Smith. In our world, Jones does not Smith.

Of course, if w is to diverge from the actual world in this way, then some miracle must occur in w. For given the laws and initial conditions that obtain in our world, Jones *must* shoot Smith at t. Thus, Jones *must* shoot Smith at t in w as well. But this 'must' can be neutralized by a miracle. (There are metaphysically possible worlds where miracles happen.) And this is just what happens in w at t. Thus w differs from the actual world in *just* the one respect the factualist requires. Consequently w can validate the counter-factual in question in just the right way.

The literature on counter-fatuals is replete with moves like the one just described. But that move fails on multiple levels. It involves easily demonstrated misunderstandings about the nature of persistence throughout time and the nature of natural law.

According to the factualist, our world is just like w *until* t: same laws, same conditions. But, in w, some miracle happens at t.

The first thing to point out is that it is not just *one* miracle that happens: it is billions upon billions. Shooting somebody is not a simple event. Various psychological, metabolic, cellular, molecular, atomic, and sub-atomic events are involved. Your pulling a trigger involves innumerable bio-chemical and, therefore, atomic events.

In w, Jones is not *supposed* to shoot Smith. He does manage to do so, but only thanks to a miracle. He is breaking the laws of nature. So each of the just described billions upon billions of atomic events are miracles. Countless particles must suddenly be re-directed or, failing that, be annihilated. This would involve suspending many, perhaps all, of the laws involved the behavior of those particles.

Let Alpha be any one of the particles composing Jones. The miracle in question involves severing many of the causal ties that link Alpha-before-t to Alpha-after-t.

Now we can begin to see the dimensions of what the factualist is demanding. Objects just *are* causal sequences. They are series of states of affairs that are linked together by physical (and, in some cases, psychological) laws. If x and y have no causal connections to each other – if they are not installments in the same causa process - then x and y are not the same object.

The factualist wants us to suspend the laws that link Alpha-before-t to be Alpha-after-t. He wants us to sever the causal connections between

the former and the latter. If he severs *all* such connections, Alpha-before-t and Alpha-after-t are siply different entities; he has simply *annihilated* a certain particle, not redirected it.

The laws that *govern* the behavior of an entity are *required* for its continued existence. Causal ties do not *bind* an object; they *compose* it. If you sever them, the object isn't liberated: it ceases to exist.

The factualist needs to sever the ties that bind Jones-before-t to Jones-after-t. But if he does this, then he has simply annihilated Jones, not redirected him. So if we grant the factualist his miracle, then *there is no Jones after t.* A fortiori, there is no Jones who feels remorse.

Of course, the factualist could argue that the miracle in question doesn't sever *all* of the ties that are involved in Jones' continued existence. Some of those ties – the ones needed to sustain Jones' existence – are left unsevered. But other ties – the ones that, in our world, lead Jones to do something other than shoot Smith – are severed.

But it is exceedingly unlikely that such a neat separation can be effected between those two sets of causal ties. *If* we are going to redirect all of those atoms, molecules, cells, and so on, then we must put all manner of physical laws out of commission. The laws of electrodynamics, mechanics, thermodynamics, and so on, must all be suspended. But if all of these laws are suspended, what is left over that can keep Jones in existence? Jones is a causal process. In all likelihood, every law of physics will have to be suspended to create the miracle that the factualst wants. (Remember the factualist is asking for *billions* of miracles; he is asking *billions* of particles to *not* do what, as a matter of physical law, they must do. Leaving aside highly derivative laws like *pinewood floats on water*, it is exceedingly ulikely that some law of physics will escape the factuliast's wrath.) The needed cut in causal ties is so deep that what is

left over, if anything, will almost certainly not be enough to sustain Jones' existence. As for the chances of Jones' continuing to exist in good enough physical and psychological *health* to feel remorse – those chances are essentially zero.

But let us leave aside considerations of physics. Let us grant the counter-factuailst all the *physical* miracles he wants. If we do this, some case can *still* made that the factualist position is bankrupt. If we consider *only* the psychological facts involved, and give the factualist all the physical facts he wants, a case can *still* be made that the miracle the factualist needs nullifies the counter-factual in question.

By supposition, up until time t, Jones' psychoogicla state in w is identical with his psychological state in our world. Of course, in our world, Jones has no intention to shoot Smith. His mind at time t is pure, at least in that regard. And the reason his mind is pure is, presumably, that in the time leading up to t, Jones either didn't have the dark desires requisite to forming an intention to shoot Smith or, if he did have them, they hadn't yet acquired the strength to crystallize in a conscious intention.

What does all this imply about Jones' mind *in w*? It implies that, in w, at time t, Jones either has no *intention* to shoot to Smith – there isn't the requisite back-ground of dark desires needed to form that intention – *or*, at time t, Jones *ex nihilo* forms an intention to shoot Smith.

In the first case, Jones (in w) is simply the victim of some kind of bodily spasm. He – or his body rather – pulls the trigger. But *Jones* is obviously not responsible. And he will probably know that. (If your body suddenly stops obeying your wishes – if you stop being able to control your arms and legs -- you will probably know it.) In that case, Jones would be *wrong* to feel remorse, and very well might *not* feel it. He would

certainly feel far less remorse than he would if he intentionally pulled the trigger.

In any case, when one says *if Jones had shot Smith, he would feel remorse*, one is not usually thinking of a case where Jones' body spasmed in a trigger-pulling way; one is a thinking of a case where Jones *intentinally* pulled the trigger. So if, in w, Jones' body spasms in a trigger-pulling way, then w will certainly *not* validate that counter-factual, at least as it is usually meant.

So let us consider the other case: the case where, in w, at time t, Jones *ex nihilo* forms an intention to shoot Smith. In that case, the intention would, by supposition, have an inadequate basis in Jones' psychological condition leading up to t. The sudden creation of this intention would constitute a discontinuity within Jones' psyche.

But it is very much an open question whether such a discontinuity is even conceptually possible. A very popular and plausible theory of personal identity is this: x and y are the same person iff some kind of *causal* connection holds between y's mental states and x's mental states. The reason I am identical with the person sitting in this chait ten minutes ago is that my mental contents are causally connected, in a certain way, to his mental contents. The discontinuity described obviously weakens the causal connection between Jones-before-t and Jones-after-t. It possibly weakens it to the point where Jones, considered as a psychological entity, simply doesn't *survive* past t.

Also, even if we waive the last consideration, it seems to me that if Jones suddenly formed an intention *ex nihilo* to shoot Smith, then his subsequent shooting of Smith would result in less guilt than if his intention had been the result of a long series of dark fantasies and ruminations. In any case, when someone says that Smith *would* feel guilt

if **he (intentionally) Smth, one is probably not assuming that the intention occured *ex nihilo* but occurred in the kind of psychological context in which such intentions usually occur.**

To sum up, the kind of miracle the factualist needs does one of the folliwng three things. (a) It makes Jones' act be unintentional. (b) It makes Jones cease to exist; it destroys the psychological continuity needed for Jones mind to continue existing. (c) It *does* allow for the existence of the evil intention; but it puts it in a sanitized psychological context – one in which it was *not* the crystallization of a hstory of dark ruminations -- where it would probably cause much weaker feelings of guilt than it would ordinarily.

On any one of these three optins, the factualist's position either falsifies or distorts the original counter-factual.

So far we have allowed that miracles *are* in some sense possible. Philosophers writing about counter-fatuals make practically unrestricted use of miracles. So do ordinary people in their day to day thinking. It is generally taken for graned that mircles are *in some sense* possible.

But the concept of a miracle is one that has no utility at all for the factualist. The very *possibility* of miracles requires that a quite specific, and highly dubious, conception of natural law be the right one. But *if* that conception of law is the right one, then the correct analysis of counter-factuals is absolutely *not* the factualist one.

Ultimately, there are two ways of thinking about laws. On the one hand, laws can be thought of as mere regularities. This is how Hume thought about them. On the other hand, they can be thought of as things that *ground* regularities – as things that *guarantee* that, when such and such happens, thus and such will also happen.

Let us consider each of these conceptions, and see what the

consequences are for the factualist.

First, suppose that Hume is wrong about laws; let us supose that laws *ground* regularities, and are not *identical* with them. I submit that, in that case, as a matter of conceptual necessity, laws are *not* the kinds of things that can be suspended.

First of all, any state of affairs that endures for more than an instant – and that may mean *any* state of affairs – simply *is* a causal process. Suspending the laws that *govern* a process means suspending the laws that allow it to exist. In effect, we saw this earlier.

The laws that govern phenomena are not imposed on them; they are internal to them.

Imagine the following. State of affairs a causes state of affairs b. The relevant covering law is: if there is an A at t, then there is a B at t*.

If a really is the cause of b, then trivially a *by itself* guarantees b's existence. It isn't as though something must be *added* to a to make b happen. If some such addition had to be made, then contrary to our supposition, a would *not* be the cause of b. It would, at most, be a partial cause.

This means that *merely* by virtue of the fact that we have a we have enough for b. In other words, given *only* that we have an instance of type A, we have enough for an instance of type B.

This, in turn, means that if the antecedent of some law is fulfilled, then *ipso facto* that law is instantiated and is operative. If the antecedent of the law if fulfilled – if we have an A – then *ipso facto* the law that generates a B is instantiated.

The only way you can have an A *without* having a B is if A is not *by itself* sufficnet for B. But if A were thus deficient, it wouldn't cause B at all. It would be, at most, a partial cause.

Clearly some states of affairs cause others. And clearly those cases of causation typically instantiate general causal laws. If one state of affairs a causes another state of affairs b, that means that a is *causally sufficient* for b. That, in turn, means that a *by itself* is sufficient for b. Thus nothing has to be *added* to a to make b happen: a is enough all by itself. This means that the operative causal law is instantiated *merely* by virtue of the fact that a exists. So once it is given that a exists, it is given that the relevant law is instantiated. And once *that* is granted, we no longer have our miracle. For the miracle just *is* the suspending – the non-instantiating – of that law.

Law are not things *added* to states of affairs; they are internal to the states of affairs they govern. You cannot remove the laws without removing the states of affairs themselves: such a removal cannot even thought, at least not coherently. (Of course, such removals *are* thought. But those thoughts involve the absurd idea that something must be *added* to a causally sufficient condition to *make* it be a causally sufficient condition.)

So if we think of laws as things that tie events together, and not just as regularities, then laws can no more be suspended or removed than triangles can have four sides.

So if the factualist is to have his miracles, then he must *not* think of laws as things that tie events together. So he must be a Humean about laws; he must see them as mere regularities.

But if laws are Humean, then the possible-worlds approach to counter-factuals is completely bankrupt.

The Humean says that laws are *just* regularities. For A's to cause B's is *just* for B's to always come right after A's. There is no inner connection; nothing *grounds* the regularity. The regularity is all that is

there. Nothing *makes* the regularity hold; it just holds.

Given this, let w and w* be two worlds that are exactly the same, in every conceivable respect, up until time t. The Humean says: there is absolutely no reason why w and w* should not diverge after t. The Humean actually allows for *any* divergence at all. The divergence can be total.

Why does the Humean allow for such a divergence? The Humean doesn't countenance the idea of there being *forces* of any kind. Nothing *makes* anything else happen. Forces, if they exist at all, are just regularities. There are no nomic necessities. There is nothing to a world *underlying* the states of affairs in it. So no matter how much two worlds are alike before t, there is no limit to how much they can differ – not just logically but *nomically* – after t.

This means that, if you are a Humean, alternative worlds are useless for validating counter-fatuals. Suppose, once again, that in our world, Bob did not shoot Smith at t. And suppose htat w is some other world where Bob *does* shoot Smith at t, but is otherwise just like our world. (Let us waive all the problems with this supposition that we've been discussing.) A Humean will say: 'I have absolutely no idea what happens in w after t. It is perfectly possible that Smith feels no remorse; it is perfectly possible that he turns into a blue-berry.' For a Humean, there will be as many worlds where Brown feels no remorse after shooting Smith – where, in fact, he turns into a blueberry or feels elated or succumbs to amnesia – as there are worlds where he does feel remorse. For a Humean, what goes on in other worlds is as likely to falsify a counter-factual as it is to negative it. Actually – depending on one's views on the arithmetic of infinitelyi large numbers -- if Hume is right, then what goes on in other worlds is *infinitely* more likely to falsify a given

counter-factual claim than it is to validate it. For every world where Brown feels remorse after shooting Smith, there are infinitely many worlds -- non-denumerably many -- where Smith doesn't feel any remorse.

To sum up, if laws are non-Humean, then they cannot be suspended. Miracles are not even *conceptually* possible. And the factualist's position fails. ON the other hand, if laws *are* Humean, then the occurrences in other worlds are at least as likely to falsify our counter-factuals as they are to validate them. So miracles are simly not things the factualists can use.

There is one incidental point I would like to make. As we've just seen, for a Humean, two worlds can be exactly alike, in every conceivable respect, before time t, and yet diverge without limit after t. This has often been taken to show that Humean laws do not support counter-factuals.

But this is absolutely the wrong conclusion to draw. The right conclusion to draw is that Humean laws don't support counter-factuals *if* counter-fatuals are seen as making statements about alternative worlds. If counter-factuals are understood in some other way, then there is a chance that Humean laws *do* support counter-factuals. We have seen, in fact, that counter-factuals *must* be understood in some other way: the alternative-worlds accont is a failoure. And we will see later that, in fact, Humean laws *do* support counter-factuals, contrary to what is generally held.

VII. Now it is time to give a positive analysis. Ultimately, I think, there is only one alternative to the analysis we've just considered. We must see counterfactuals as making purely *epistemic* statements – as making statements about relations between *propositions*: they must be statements to

the effect that one proposition gives a very high degree of probability (perhaps 100%) to some other proposition.

Let us get clear on a couple of basic points about probability. First of all, probability is typically a relation between *propositions*, not states of affairs. Statements like the following are often true:

(i) The probability that Socrates had four kidneys is 5%.

But either Socrates did, or he did not, have four kidneys. It couldn't be 5% true that he had kidneys. The same is true *mutatis mutandis* of

(ii) The probability that Bill is on the North Pole right now is .002%

But obviously it cannot .002% true that Bill is on the North Pole. It is either a 100% true or 0% true.

Obviously (i) and (ii) make statements about the credibility that certain propositions give to other propositions.[105] Let us focus on (i). Anyone affirming (i) is affirming it on the basis of certain pieces of (putative) knowledge that he has. He knows that most people don't have four kidneys, but that some people do. He knows, on the basis of a careful reading of the platonic dialogues, that Socrates had some of the symptoms of the symptoms had by people with that condition have. But he also knows that Socrates had some other illness which would explain his having those symptoms. And so on. Let $P_1 \ldots P_n$ be the propositions believed in. In affirming (i), our man is really saying that $P_1 \ldots P_n$ give a certain degree of credibility to the proposition that Socrates had four kidneys. Obviously the same considerations (mutatis mutandis) apply to (ii)

So (i) – or, rather, any token of it—is really elliptical for something of the

form: $P_1 \ldots P_n$ *give such and such credibility to the proposition that Socrates had four kidneys*, where $P_1 \ldots P_n$ are propositions believed true (or probable).

Of course, different tokens of a probability-sentence can be made on the basis of different data. When this happens, different statements are, strictly speaking, being made. Joe might token (i) on the basis of some piece of knowledge unknown to Jerry, who tokens the *same* sentence. We must hold that, in such a case, Joe and Jerry are really making different statements: Joe is saying

(iii) *S gives such and such credibility to the proposition* Socrates is bald

where S is some conjunction of proposition.

On the other hand, Jerry is saying

(iv) S* gives *such and such credibility to the proposition* Socrates is bald

where S* is some set of propositions other than S.

It is clear, then, that probability is (typically) an *epistemic*, not a *metaphysical*, notion. When you utter (ii), you are saying that, given the data at your disposal, it is *epistemically* possible to degree .05 that Socrates will turn out to have had four kidneys. Obviously it is not *now* metaphysically possible for Socrates to have had four kidneys (in this world) unless he *actually* had four kidneys: *that* is an all or nothing affair – no percentages can be attached to a *truth.* So when you say (i), you couldn't possibly be talking about what is now metaphysically possible, but rather about what is now epistemically possible, which means: what is not ruled out by the propositions that are currently known to you.

I have not gone so far as to say that probability is *always* a propositional or

(what is nearly enough the same thing) an epistemic notion. I allow that there may be a non-propositional, non-epistemic notion of probability.[106] If it really is nomically undetermined when electron e will jump to orbit O, then perhaps the proposition 'the probability that e will jump to O at t is .004% ' has a non-propositional reading. (It also has a propositional reading: proposition $P_1 \ldots P_n$ give probability .004 to the proposition *e jump to O at t* – where $P_1 \ldots P_n$ are, presumably, propositions about the behavior of similar electrons in the past, about physical laws, and so on.) But it is beyond dispute that probability has *an* epistemic/proositional meaning. And that is we need to know for our purposes.

The meaning of

(v) if Kennedy hadn't died in '63, Johnson wouldn't have become president in '63

can only be

(vi) The propositions $P_1 \ldots P_n$ give a high degree of credibility to the proposition *Johnson didn't become president in '63*,

where, for some i, P_i is the proposition *Kennedy didn't die in '63.*

It must be kept in mind that *false* propositions probabilify other propositions just like true ones. The proposition *the moon is made of cheese* is false, but that doesn't stop it from giving a very high degree of probability (essentially 100%) to the proposition *there is a piece of cheese that weighs more than 100 lbs.* The proposition *Kennedy didn't die in '63* is false, but it still gives probability to other propositions. It gives nearly 100% probability to the proposition *some living organism weighing over 10 lbs didn't die in*

'63. And it gives some probability, though not nearly as much (at least not when taken by itself), to the proposition *Johnson didn't become president in '63.*

Notice that our purely propositional analysis of counterfactuals doesn't have the problem faced by the factualist analysis. The factualist analysis needs to keep the laws the same but vary the events: which, as we've seen, it is typically not *logically* possible to do without undermining the counterfactual. Given that it is *logically* impossible, it follows that it is *metaphysically* impossible. (Anything which is logically impossible is metaphysically impossible, though not vice versa.) So the factualist needs the *metaphysical* possibility of something which isn't metaphysically possible. But *we* don't need the metaphysical possibility of *anything.* What *we* do is merely to register the confirmation relations of some proposition, namely *Kennedy did not die in '63.* We don't need there to be a *world* that has the same laws as ours but where Kennedy didn't die. Such a world may well *not* be possible, in any but an epistemic sense. For if the laws in w are the same as ours, then any difference between our world and w that is sufficient for keeping Kennedy from dying is almost certainly going to be sufficient for a lot of other things that will make (v) be false – it might be sufficient for there not being a Kennedy or for Kennedy's dying of hemorrhagic fever in '63. All *we* need to make (v) true is the existence of some set of propositions $P_1 \ldots P_n$ – where *Kennedy did not die* is among them – such that those propositions give a high degree of credibility to the proposition *Johnson did not become president in '63.*

There is, I grant, a bit more to it than that. (To facilitate discussion, let us suppose that P_n is the proposition *Kennedy did not die in '63.*) We cannot, of course, allow just *any* proposition to be included in $P_1 \ldots P_{n-1}$. It isn't clear what *all* of the selection criteria are, but it is clear what *some* of them must

be. Trivially, $P_1 \ldots P_{n-1}$ cannot include the proposition *Johnson didn't become president in '63.* Less trivially, $P_1 \ldots P_{n-1}$ cannot include propositions *known* to be false; for example, P_3 cannot be the proposition *the black plague kills all American politicians (including Johnson) in '63.* Also, $P_1 \ldots P_{n-1}$ cannot *ex*clude propositions that we know to be true. Suppose we knew that Kennedy was dying of a rapidly progressing form of cancer which kills its victims in under three weeks. Obviously we couldn't exclude that from $P_1 \ldots P_{n-1}$. The rule seems to be that $P_1 \ldots P_{n-1}$ must include all and only known propositions.

But perhaps this rule is a bit simplistic. What a biologist considers counterfactually possible differs from what a physicist would consider counterfactually possible. If you ask a biologist whether a given zygote *could* have acquired a certain morphology in due course (say, that of a male), he may say 'why of course: it is completely open which gender a zygote assumes – so of course it *could* have acquired that morphology'. But if you ask a physicist the same question he may say: 'given the initial positions and momenta of the atoms composing that zygote and composing the neighboring physical systems [and so on]…it was inevitable that it would eventually assume that morphology.' The biologist seems to be saying: nothing that falls within the scope of *biology* rules out the possibility that the zygote would become a female, not that nothing *tout court* rules it out. The physics is saying: something that falls within the scope of *physics* rules out the zygote's becoming a female.

Similarly, if you ask an economist and a historian what was counterfactually possible – whether Saddam *would* have done such and such *if* thus and such – they may give different answers; for the historian will relativize his statement to propositions falling within the scope of his broadly historical concerns, whereas the economist will consider a slightly different data set.

But this relevance constraint might well reduce to the fact that the physicist *knows* certain things that the biologist does not and that the economist *knows* certain things that the historian does not (and perhaps vice versa). So it may be true that the one selection criterion is: $P_1 \ldots P_{n-1}$ must include all and only those propositions known (or reasonably believed) to be true by the person affirming the counterfactual (excluding, of course, the proposition whose negation is antecedent of the counterfactual).

And it is clear that, given any one person, a great deal is epistemically possible from his perspective that is not *objectively* (metaphysically) possible. It was epistemically, but not metaphysically, possible that Hesperus should be distinct from Hesperus. Given only what *I* know, it is epistemically possible that the laws of physics remain exactly what they are *and* that there will not be an earthquake in California in the next year. But it might well be a *logical* impossibility – not a nomic or even a mere metaphysical impossibility – that the conjunction:

$P_1 \ldots P_n$ and $C_1 \ldots C_n$ and *there is no earthquake in California in the next year*

should be true, where $P_1 \ldots P_n$ is the set of propositions giving the laws of physics and $C_1 \ldots C_n$ is the set of propositions describing the current condition of the California landmass.

Basically, for *our* analysis of counterfactuals to work, it is necessary only that *relative to what is known*, some untrue proposition, conjoined with propositions embodying the just mentioned knowledge, give high credibility to some other proposition. We don't need some set of initial conditions that *didn't* hold to *actually* hold, albeit in some other world, all the while making sure that *our* laws apply in that other world. For all *those* desiderata cannot be

jointly satisfied. All we need is for some proposition (the consequent of the counterfactual) to be epistemically possible *relative to some false proposition* (the antecedent of the counterfactual) conjoined with *what we know*. And that is an extremely liberal constraint. For what is possible relative to what we know is extremely inclusive. And it *is* possible, relative to what I know (minus my knowledge that Johnson *did* become president in '63) along with the proposition *Kennedy didn't die in '63* -- that Johnson didn't become president in '63. In other words, if $p_1 \ldots p_n$ are the propositions embodying what I know, minus the proposition that Johnson became President in '63, then the conjunction of $p_1 \ldots p_n$ and *Kennedy didn't die in '63* gives high probability to the proposition *Johnson didn't become president in '63*. And that is what it is for the counter-factual *if Kennedy didn't die in '63, Johnson wouldn't have become President* to be true.

Now if I were an extremely knowledgeable historian, it might *not* be the case that, relative to what I know, *Johnson didn't become president* was epistemically possible for me. But that is consistent with the nature of counterfactuals. The more a person knows, the fewer counterfactuals he can (coherently) concede or even entertain. If you ask somebody who knows nothing about American history what was *possible* for the country in 1813, he will say 'just about anything.' If you ask somebody who knew exactly how much money was in the treasury, exactly what the President at the time was thinking, exactly where the troops were stationed, and so on, that person will probably have a different answer. This point about possibility obviously maps onto one about counterfactuals: assuming maximal rationality on the part of the respondents (so that we don't have to worry about incompetence on their part)[107], the question 'what *would* have happened to the U.S. if…?' gets very different answers from different people depending on what those people know. This doesn't mean that one of them is right and the other is wrong. It

means that the question is inherently epistemic. One of the respondents, because he knows too much, cannot (without contradicting himself) fit the antecedent of the counterfactual into the list of propositions he knows to be true. The antecedent isn't epistemically possible for him (given what he knows). So he rightly balks at the question. But the antecedent *is* epistemically possible for the *other* respondent; so he doesn't balk.

Notice that if you speak to somebody who is extremely knowledgeable about some domain, it is very hard to phrase 'what if…?' questions concerning that area that he will not dismiss as senseless; he will almost always say 'well, that isn't an easy hypothetical to run because Saddam was already in the process of…' or 'well, if you put a thermometer in such a liquid then…'. The more a person knows, the less is epistemically possible from his perspective, and the less likely he is to be able to insert the antecedent of a counterfactual into his existing knowledge-base. So I think that our epistemic analysis of counterfactuals has some support, at least on that level.

VII. **There are many apparent problems with this analysis. I'd like to say why, in my view, they are not actual problems for it.**

According to your analysis, the meaning of "if A were the case, then B would be the case" is:

P(A) (together with certain other propositions) gives a high-degree of credibility to P(B),

where P(A) is the proposition affirming A's existence, and P(B) is the proposition affirming B's existence.

It seems to me that your analysis is vulnerable to an exact analogue of the argument you directed against the factualist analysis. If A were the case, it would obviously be caused by some pre-existing state of affairs C_1. And C_1, in its turn, would be caused by some pre-existing state of affairs C_2. And so on. Of course, for each i, C_i would be a state of affairs that did not *actually* obtain. For reasons you described at great length, the conjunction *P(A) and* $P(C_1)\ldots P(C_n)$ would either give little or no probability to P(B) *or* that conjunction would probability propositions that were not even counter-factually true. So your analysis fails for the very same reasons that the factualist analysis fails.

This point overlooks the very thing that distinguishes my epistemic analysis from the factualist analysis. Imagine the following. You and I are homicide detectives. We are trying to solve the murder of Smith. We did not witness the murder, and neither did anyone else. But we have data that incriminates Brown. Brown's finger-prints are found on the knife lodged in Smith's rib-cage. Specimens of Smith's hair and blood were found in Brown's car and home. And so on. Let $P_1\ldots P_n$ be the propositions that describe all of the data that we've gather so far, along with the propositions that embody our knowledge of regularities and causal laws. And let P(M) be the proposition: *Brown killed Smith.*

***By themselves,* $P_1\ldots P_n$ give a certain degree of probability to P(M). That probability is less than 100%, but is still quite high. Given *only* $P_1\ldots P_n$, it is still open whether Brown murdered Smith. $P_1\ldots P_n$ don't *entail* P(M). Perhaps Brown is being framed, and the evidence in question was planted. Perhaps, unknown to the world, Brown has an identical twin – who has the same finger-prints and genetic profile as**

Brown – and that twin committed the murder. That is all well and good. But there is no disputing that $P_1…P_n$, *by themselves*, give a certain probability to P(M). We don't need *new* data to decide whether $P_1…P_n$ probability P(M). They obviously do probabilify it.

Of course, there are many propositions P_{n+1} such that $P_1…P_{n+1}$ give extremely low probability to P(M). P_{n+1} could be the proposition: *Jones, not Brown, killed Smith.* This would be a trivial example. A less trivial example would be the case where P_{n+1} is the proposition: *Brown has a sociopathic identical twin who has the same genetic markers as Brown and also has a key to Brown's car and house.*

But the proposition *$P_1…P_n$ give 95% probability to P(M)* is simply distinct from *$P_1…P_{n+1}$ give 95% probability to P(M).* In fact, those two propositions *must* have different truth-values, unless P_{n+1} is some completely irrelevant datum.

Of course, we do need new data. But we need new data *not* to decide whether $P_1…P_n$ probabilify P(B). We already know that they do probabilify it. We need new data to decide whether *in fact* P(M) is true. But it is beyond dispute that, by themselves, $P_1…P_n$ probability P(M).

Indeed, new empirical data will not have any bearing on *how much* $P_1…P_n$ probability P(M). New empirical data will bear on whether P(M) is true or not. But the relationship between $P_1…P_n$ and P(M) is a purely logical relationship. Granted $P_1…P_n$ do not *entail* P(M). But the proposition: *if $P_1…P_n$, then P(M) has a 94% likelihood of being true* is not itself an empirical proposition. *$P_1…P_n$* is an empirical proposition. And so is P(M). But the conditional proposition *if $P_1…P_n$, then P(M) has a 94% likelihood of being true* is not empirical. (This crucial point about probability statements is made and decisively argued for in

Bonjour (1998).)

When one makes a probability-assertion, one is making a statement about the bearing of what one *knows* – not what *is* the case -- on some proposition not yet known to be true. One is *not* making a statement abou the bearing of the *totality* of true propositions on some other porpositon. Socrates either had two kidneys or he didn't. And yet statement *it is 98% likely that Socrates had two kidneys* may well be true. That statement does *not* mean: *the conjunction of all true propositions gives 98% probagility to the propostion that Socrates had two kidneys.* Probability statements *never* have that form. For the conjunction of all *true* propositions will either give 100% or 0% probability to any given proposition, depending on whether the given proposition is true or false. And, outside of some very special contexts, we are not interested in such cases. Probability statements have the form

(*) *if P, then the probability of Q being true is such and such.*

If we *add* propositions to the antecedent, we get an entirely different probability statement. If, say, we *add* $P_2 \ldots P_{37}$ to the antecedent, then we end up with the statement

(**) *if P and P_2 and…P_{37}, then the probability of Q is such and such.*

But (*) is a completely different proposition from (**). If we *add* material to the antecedent of a probability claim, we simply get a new probability claim. So it is actually *obligatory* not to add new such information. In particular, it is obligatory *not* to add information having to do with alleged causal-preconditions for the truth of the antecedent. If we did

add such information, then we would simply be considering a different probability claim.

My analysis of counter-factuals is this: *if A had been the case, then B would have been the case* means this: The proposition

(***) *K and P(A) gives a high probability to P(B),*

where K is the conjunction of proposition we know to be true, minus the proposition *A was not the case.* Of course, (***) is equivalent to:

(****) *if K and P(A) are the case, there it is highly probable that P(B) is true.*

If we *add* propositions to the antecedent of (****), then we simply end up with a *different* proposition from (****). The objector says that we must add such antecedents. But the truth is: we must *not* to do so. (And this obligation *not* to do so lies at the heart of the concept of probability.) So there is no regress of any kind, and no analogue of the argument we used against the factualist is generated.

VII. Here is another objection to our analysis:

An alternative world is usually thought of as a set of propositions. Let's suppose this is right. In that case, I don't see how your view differs from that of the factualist. You say that a counter-factual registers some relation – presumably one of entailment or confirmation – holding among prpositions. The factualist says that a counter-factual says what is the case in some other world. If that other world just *is* a set of prpositoins, then he is really

affirming the existence of some kind of relationship -- presumably one of entailment or confirmation -- among propositions. He is saying that, if P1… Pn are true, then so (probably) is some other proposition Q. (Q is the consequent of the counter-fatual. Among P1…Pn is the proposition affirming the antecedent of the counter-factual. The remainder are those propositions that describe our world *minus* the ones that are inconsistent with the proposition affirming the antecedent of the counter-factual.)

Basically, if worlds just *are* sets of proositions, then the factualist approach collapses into your epistemic approach. And your approach therefore cannot possibly be superior to the factualist approach.

There is a certain truth in this comment. *If* altenrative worlds are mere sets of propositons, then the factualist one isn't so different from the epistemic one I advocatge.

But there are two problems with the objector's point. First, worlds are not sets of propositions. Second, while it is true that many factualists identify worlds with sets of propositions, the sets they have in mind – so-called "maximal consistent sets" -- are absolutely useless for analyzing counter-factuals.

When I use the term "world", I mean what is ordinarily meant by that term: I mean the kind of thing that *this* world is. Our world is a good example of a world. It is, in fact, the only uncontroversial example we have. Our world is not a set of propositons. Our world is something which makes sets of propositions be true or false. Any world deserving of the name is something which makes propostions be true or false. A set of propositonis does not make propositions be true or false. The set of propositonis <snow is white, grass is green, penguins fly> doesn't make any of those porpositnos be true

or false. That set would exist even if grass were orange. That set exists regardless of the color of snow or grass and regardless of the flying capabilities of penguins. Any world deserving of the name is not a set of propositions.

Granted, there are may who maintain *both* that counter-fatuals register facts about other "worlds" *and* say that other "worlds" are just sets of prposiitns. Let us refer to such people as "weak factualists". (A "strong factualist" would be somebody, like David Lewis, who is a factualist and who holds that alternative worlds are concrete entities, not sets of propositions.) There are many weak factualists. In fact, so far as I know, *every* one who writes about counter-factuals, apart from Lewis, is a weak factualist. But weak factualists never, absolutely never, hold that counter-factuals register relations of probabilification among propositions. In my view, that is what they *should* hold. But it isn't what they hold.

What *do* weak factualists hold? It seems to me that what is going on is this. They start out by accepting Lewis' analysis. But they are not willing to go so far as to say that there really are alternative worlds, in any literal sense of "world". But they want to hold onto a Lewisian analysis – because that analysis obviously does an excellent job of capturing our pre-theoretical understanding of counter-factuals. So, on the one hand, the weak-factualist isn't willing to completely give in and treat counter-fatuals as inter-propositional relations. (That is why, prior to this paper, no one has ever said flat out that *if Kennedy had not been shot, Johnson would not become President in '63* holds entirely in virtue of some relationship of probabilification holding among propositons.) But on the other hand, the weak-factualist isn't willing to go in the other direction: he isn't willing to say that *there really are* other worlds. So the weak-factualist denies that counter-factuals mrely register relations of probabilification among

propositnos. *And* he denies that counter-factuals register facts relating to other worlds (in any genuine sense of “world”). So his position is completely incoherent.

If we are to understand counter-factuals, we must completely jettison the notion of alternative world. If by “world” we mean something at all ike our world, then counter-factuals are not to be understood in terms of worlds. If by “world” we just mean “set of propositions”, then we should just talk about sets of propositions, and not let a misleading bit of terminology create the illusion of explaantion.

There is more to say. Apart from Lewis, everyone who uses the notion of other worlds world idenifies worlds with sets of propositions. But they don’t ever say that a world is just *any* set of propositions. They always identify worlds with what are called *maximal consistent sets*. What is a maximal consistent set of propositions? S is a maximal consistent set of propositions iff, for any atomic proposition P[108], either P or its negation is included in S.

Why are worlds identified with *maximal consistent* sets? The reason is clear. Consider the following set of propositions. <Grass is green, penguins fly>. This set of propositions obviously doesn’t describe a world; much too much is left out. Any number of *different* worlds could accommodate this set. A world is supposed to something which is determinate in every respect. The set <Grass is green, penguins fly> leaves too many things indeterminate. Suppose, for the sake of argument, that that set *were* a world – call it w. In that case, the following questions would have no answer. “Are there tigers in w? Are there nine planets in the solar system in w? Did Keats live to be 90 in w?” Obviously w is much too thread-bare to count as a world: a world is something which gives answer to questions.

So if worlds are, with any plausibility, to be identified with sets of

propositions, they must be sets of propositions that don't leave too many questions unanswered. To turn w into something that is even minimally deserving of the title "world", we have to flesh it out – we have keep adding propositions to it until w becomes capable of settling all the questions we might pose about it. This means that, for any question Q we might have about w, one of the propositions in w must be an answer to Q *or* the propositions that are in w must *entail* an answer to Q.

Now we can see why alternative worlds, in the sense of maximal consistent sets of propositions, are absolutely useless for analyzing counter-factuals. Consider the counter-factual

(JK) *if Kennedy had not been assassinated, then he would have been president until '68* .

Here is what the weak-factualist says.

(KJ) is true because in some world w, Kennedy was not assassinated and, in w, he remained president until '68. w is a set of propositions. w comprises the proposition *Kennedy was not assassinated.* With that qualification, w is as much like the set of propositions that describe *our* world as any consistent set of propositions could be. w also comprises the proposition *Kennedy was president until '68.*

But, in effect, we've already seen the problem with this view. If w is to comprise the proposition *Kennedy was not assassinated*, then either w is will be replete with inconsistencies *or* w will describe a world so utterly different from our world that it almost certainly does not comprise the proposition *Kennedy was president until '68.*

Remember what we said about the strong-factualist position. If w is a world – in the literal sense of "world" -- in which Kennedy is not assassinated, then w will differ from our world in so many ways that Johnson could very well be president in '63. This fact about w corresponds to a point about the set of propositions *describing* w. That set must include *Kennedy was not assassinated.* But if that set is to be consistent, then it must be dramatically different from the set of propositions which describes our world. But if the former set is *that* different from the latter, then it is, at best, minimally likely that it also comprises the proposition *Kennedy was president until '68.*

All the points that we earlier made about worlds are really points about propositions describing worlds. The position of the weak-factualist is very close to that of the strong-factualist. The only difference is that the weak-factualist talks about the propositions *decribing* worlds, whereas the strong-factualist talks about worlds themsvles. But all of the problems bedeviling the srong-factualist positon are problems with the proposition *describing* worlds: there is no way to insert the antecedent of a counter-facctual into that set without making the set prohibitively different from the set of propositions which dexcribes our world. So the weak-factualist position is barely any better than the strong-factualist one.

The problem with the weak-factualist position is that maximal consistent sets comprise *too many* propositions. A maximal consistent set answers *too* many questions. Among those questions are questions relating to nomic relations among propositions. So if a consistent set is going to comprise a proposition affirming some proposition that is not true of our world – some proposition that could be the antecedent of a counter-factual – then either (a) that set wll describe a world that bears no resemblance to ours *or* (b)that is going to have to leave a lot of questions unanswered. If (a), then the set is useless for running counte-factuals. If (b), then the set is not a maximally

consistent set, but only some tiny subset of such a set. So the weak-factualist ends up identifying worlds with tiny subsets of maximal consistent sets of prospotisions.

I grant that *if* he does this, then his position can work. But if he does that, then his position does not have anything to do with "worlds" at all, not even relative to the most broad and inclusive intepretatin of htat term. Everyone who talks about possible worlds is either talking about concrete worlds or about *maximal* sets of propositios. On either reading, the correct analysis of counter-fatuals cannot possibly have anything to do with possible worlds.

Given this, it is easy to deal with another objection I have encountered:

Your analysis is supposed to *not* use the notion of a possible world. But it *does* in fact use that notion. You analyse counter-factuals in terms of the notion of epistemic possibility. According to Hintikka, a prpositoin P is epistemically possible for so and so iff P is true in some world that is accessible to so and so. This is a standard view.

Your analysis is roughly this. *If S had been the case, then S* would have been the case* means that P(S), conjoined with the propostions that are epistemically possible for you, entail P(S*), where P(S) and P(S*) are the propositions afriming the existence of S and S*. When we apply Hintikka's conception of epistemic possibility to this, what we get is: *If S had been the case, then S* would have been the case* means that, in every world epistemically available to you where S is the case, S* is also the case. So your anlaysis collapses into some version of the factualist analysis.

First of all, Hintikka is not giving a *definition*: he is giving an *analysis.* The correct, and uncontroversial, *definition* of "epistemic possibility" is this. A

proposition is “epistemically possible” for me if, given what I know, I can believe it to be true without being guilty of self-contradiction or, more generally, incoherence. If Socrates was human, then presumably he was *necessarily* human. But it is epistemically possible for me that Socrates was not a human, and was actually a martian. For given only what I know, I would not be guilty of self-contradiction if I were to believe that Socrates was a martian.

Hintikka is giving us an *analysis* of the notion of epismite possibility, i.e. of when a prpositoin can coherently be believed. He says: X can coherently believe P, i.e. P is epistemically possible for X, just in case P is true in some world that is accessible to X.

Now *if* Hintikka’s analysis of the concept of coherent/epistemic possibility is correct, then perhaps my analysis collapses into some version of the factualist one. But Hintikka’s analysis is not correct; it fails in two respects.

By a “world”, Hintikka either means something concrete – something like *this* world – or he means some kind of set or propositions. Let us consider each option. First let us suppose he means smoeting concrete. In that case, his anlaysis is simply false. For it is well known that what is epistemically possible may not be *metaphysically* possible. It is epistemically possible that water is not H2O. But there couldn’t possibly be any world (in the sense of a concrete situation) where water is anyting other than H2O. *A fortiori* there are no *available* worlds where water is not H2O. So if, by “world”, Hintikka means some kind of concrete object, then his analysis is simply false: it doesn’t allow that what is epistemically possible should diverge from what is metaphysically possible.

So by “world”, Hintikka must *not* mean some kind of concrete object; therefore, he must mean some kind of set of propositions. In that case, his analysis becomes: P is epistemically available to X iff P is true in some world

that is available to X, and a “world” is a set of a propositios. In other words:

P is epistemically available to X iff P is true in some set of a propositions that is available to X.

To evaluate this proposal, we must first answer to questions. Frist, what is it for a porpositoin P to be true “in” some set of propositons? Second, what is it for a set of prositons to be “available” to someone?

Presumably, P is true “in” some set of propositions $P_1 \ldots P_n$ iff the truth of P is required for the truth of the conjunction of $P_1 \ldots P_n$. So “snow is green” is true in the set of propostions <grass is purple, snow is green, penguins play gold> because, if *snow is green* were false, then the conjunction of those three propositions would be false.

Now *any* proposition, no matter how absurd, is true “in” some set of propositons. Consider the set <1+1=3, penguins play golf, triangles have three sides>. The proostion *1+1=3* is true in that set – in that “world” or world-fragment.

Presumably, a set of propositions is “available” to X if X grasps, or can grasp, those propositions. A single proposition – a set of one proositoin – is “available” to X iff X grasps, or is able to grasp, that proposition.

Now we can see the problem with Hintikka’s view. 1+1=3 is definitely *not* epistemically possible. But anyone can grasp that proposition. And anyone can grasps *sets* of propositions containing that proposition. So “worlds” containing that proposition *are* “available” to people. Thus Hintikka’s analysis predicts that *1+1=3* is epistemically possible. But it isn’t.

Hintikka has an obvious response. He can say:

When I say that P is “available” to X, I don’t mean merely that X can grasp P. After all, people can obviously grasp all kinds of absurd propositions. I mean that X can *coherently* believe P. So P is epistemically possible for X iff P is “available” to X. And P is “available” to X iff X can *coherently* believe P. So, despite what you just said, my analysis doesn’t have the absurd consequence that *1+1=3* is epistemically possible.

But now Hintikka’s anlayiss is, at best, viciously circular and, at worse, completely devoid of content. Hintikka is trying to *analyse* the notion epistemic possibility/coherence in terms of the notin of available worlds. So when we ask Hintikka what he means by “available world”, he cannot, without vicious circularity, say “an available world is a set of propositions that can coherently be believed”.

VIII. These points enable us to deal with another objection:

You say that counter-factuals register relations of some kind holding among propositons, not among worlds. These days propositions are generally held to be sets of worlds or, at any rate, functions involving worlds. If this conception is the right one, then I don’t see the difference between your “epitemic” analysis and the traditional “factualist” one. Both analyses, apparently, turn counter-factuals into relations holding among proopsitons.

Again, we must distinguish definitions from analyses. It is, indeed, widely held that propositonis are sets of worlds, or functions invovign worlds. But this is not a definition of “proposition”; it is a controversial *analysis* of what propositions are.

The uncontroversial *definition* of “proposition” is this: propositions are

those things that are non-derivatively true or false. Sentences are true or false, but only because they bear propositions. (“Snow is white” is true entirely because it bears the proposition *snow is white*.) Beliefs are true or false, but only because they encode prpositions. (My belief that snow is white is true entirely because it encodes the propositoni *snow is white*.) Whenever we have something that is true or flase, we *either* have a proposition *or* we hve something which inherits its truth-value from a proposition.

So, in the end, prpositions are the things that are *non-derivatively* true or false. This is fairly uncontroversial. And when I use the term “proposition”, all I mean is: something which is non-derivatively true or false.

The objector says that *if* propositions are sets of worlds, or functions involving worlds, then my analysis is not (appreciably) different from the factualist one. I must disagree.

If someone analyses proositoins are sets of worlds, or as functions from worlds to truth-values, he cannot then say that worlds are themselves sets of proostions. For then his analysis of “prposition” is this: a propositoniis a set of sets of propositions or is a function from sets of sts of propositions to truth-values. Such an analysis would obviously be circular in the worst way. To sum up, someone who analyses propositons as sets of worlds, or functions involving worlds, is quite definitely committed to the view that worlds are *not* sets of propositions; he is therefore committed to the view that worlds are concrete objects.

As we have seen, if worlds are concrete objects, then the factualist position fails. So contrary to what the objector says, the factualist must hold that prpositonis are *not* sets of worlds or functions involving worlds.

But even if we leave aside the last point, the objector’s point fails. For the sake of argument, suppose the objector is right about propostions: suppose that, indeed, propositions are sets of worlds or functions involving worlds.

Within this framework, how are to anlayse the notion of entailment? In other words, supposing that propositions are sets of worlds or functions involving worlds, what is it for P to entail Q?

The only answer that is possible within this framework is: P entails Q iff every world in which P is true is a world in which Q is true.[109] In other words, Q is the case in every world in which P is the case.

But this answer is not the right one. *Water is wet* is true in every world in which H_2O *is wet* is true, and *vice versa*. But no entailment relation holds between the two propositons. A relation of necessitation holds, but not one of *entailment.*

Consier the proposition: *It is not the case both water is wet and H2O is not wet.* One can believe this without being guilty of incoherence. Thereofer *water is wet* does not entail *H2O is wet.*

The relation entailment is a special case of the relatin of necessitation. If P entails Q, it follows P necessitates Q, but not *vice versa.* Pws says that P necessitates Q iff *not [P and not Q]* hoilds in *every* world. So pws must say that P *entails* Q iff *not [P and not Q]* holds in every member of a certain *subset* of the set of all worlds.

There are no worlds in which the following is true: x is water and is not H2O. That is why *x is water* necessitates *x is H2O.*

There are no worlds in which x is a triangle and is does not have three sides. According to pws, that is why *x is a triangle* necessitates *x has three sides.*

But *x is a triangle* not only necessitates *x has three sides.* The first *entails* the second. By contrast, *x is water* does not entail s *x is H2O.*

How does pws explain this difference? Pws has only a couple of options. It can say: there are *epistemically* possible worlds where the proposition [*x is water and x is not H2O*] is true.

But this is doubly incoherent. First of all, there are *no* worlds where that

proposition is true. *A fortioroi* there are no epistemically possible worlds where it si true. The idea that there are "worlds" that are epistemically, but not *metaphysically*, possible is simply absurd. If something is metaphysically imposible, that means it doesn't exist. Pws inadvertently collapses epistemic possibility into metaphyhsically possibility.

Second, as we saw earlier, the notion of epistemic possibility is to be defined in terms of the notin of self-contradiction. So pws canot use that notion to analyse the notion of entailment or, therefore, of self-cnotradiction.

What other options does pws have? It could say this:

there are no worlds where *[x is water and x is not h2O is true*]. And there are no worlds where *[x is a triangle and x does not hav three sides]* is true. Nonetheless, one isn't guilty of incoherence if one believes that there are worlds of the former kind. But one *is* guilty of incoherence if one believes in worlds of the second of kind. That is why *x is a triangle* entails *x has three sides*, whereas *x is water* does not entail *x has three sides*.

But if he says this, the proponent of pws is giving up on the idea that entailment can be analyzed in terms of relations among worlds. For he is ,in effect, saying that the relevant relation among worlds must *itself* be understood in terms of entailment.

Uncontroversially, P entails Q iff [P and not Q] is self-contradictory. The objector says: P entails Q iff it is self-contradictory to believe in a world where [P and not Q]. This is tantamount to: [P and not Q] is self-contradictory iff it is self-contradictory to believe in a world where [P and not Q].

This last point is true enough. But it is no *anlaysis* of the notin fo self-cnotrdiction. It *uses* that notion; it *presupposes* it. It is therefore no analysis

of it.

Pws completely and utterly fails to elucidate the notion of entailment. Whenever we try to anlayse entailment in terms fo relations among worlds, we end up presupposing that very ntoin . The notin of entailment is intimately related to the notion of a propsotiion. In fact, a standard view is that a prooption P *is* the class of its logical consequences.[110] The rleatnioship between the notion of entailment and propositionality is *so* deep that, if a doctrine fails to anlayse the former, then it surely fails to analyse the latter. Pws fails to anlase the former. We msut conclude that it fails to anlayse the latter: it doesn't give a correct of what propostions are. Propositions are *not* sets of worlds or functions involving worlds. The notion of a proposition is *not* to be anlaysed in terms of the notin of a world.

VIII. I would like to bring up one consideration that favors our analysis. The factualist analysis of c counterfactuals is, characterized by a certain *prima facie* circularity. There are ways that the factualist can break the circle. But those ways actually make matters worse for the factualist.

Not all true universal statements express laws. Consider the statement: *Every coin in my pocket is a quarter*. That is a true universal statement.For it is equivalent to

(*) for any x in the universe, if x is a coin in my pocket, then x is a quarter. [111]

But (*) does not express a law. Let us contrast (*) with a proposition which *does* express a law (albeit a highly derivative one):

(**) Pinewood floats on water.

What is the difference between (*) and (**)? The generally accepted answer is this: (**) supports counter-factuals; (*) does not. If I *were* to put a nickel or a penny in my pocket, it wouldn't turn into a quarter. So the counter-factual claim

(***) For any x, if x *were* a coin in my pocket, it would be a quarter

is false. And that is why (*) does not express a *law.* By contrast, the counter-factual claim:

(****) For any x, if x were a piece of pine-wood placed on water, then x would float

is *true.* And that is why it expresses a law.

So what is a law? The answer most would accept is:

(L) A law is what is given by a universal proposition that supports counter-factuals.

Here is the problem for the factualist. The factualist says that counter-factuals make statements about other worlds that are like this world, except in some very limited respect. When we ask the factualist "how are those worlds

like our world?", he says (among other things): 'well, those other worlds are governed by the same *laws* as ours.' So the counter-factualist says:

(CF) A counter-factual makes a statement that concerns (or is made true by) facts in worlds that are governed by the same laws that apply in our world.

Very few deny (L). In fact, practically *everybody* – factualists included[112] – believes (L). So when we ask a factualist what a law of nature is, he says "it is something that supports counter-factuals". And when we ask him what a counter-factual is, he says "something that is made true by worlds that have the same laws as our world". This, it seems to me, is plainly circular.

Granted, not all circularity is vicious. I believe both of the following propositions: (a) bachelors are unmarried males and (b) unmarried males are bachelors. But I am not, on that account, guilty of circularity. All tautologous beliefs, possibly all analytic beliefs, have an element of circularity. But one is guilty of vicious circularity in virtue of having such beliefs.

Before we can see clearly why the factualist is guilty of vicious circularity, there is one more nuance we must deal with. In contemporary philosophy of science, it si a common-place[113] that scientific theories are characterized by a *certain* circularity. A theory is a set of inter-connected, mutually supporting propositions. OF course, a theory has to have *some* roots in data lying outside that theory. But *within* a theory we may find many cases of circularity. (Facts about protons will explain facts about electrons, and *vice versa*.) The theory *as a whole* cannot be circular: it has to be anchored in, and answerable to, extra-theoretical facts. But *within* the theory, there are many cases of prpositions giving explanatory support to *one another*. So, once again, we see that not all circularity is vicious.

This is fine and good. But it is irrelevant to the factualist's situation. Scientific theories are causal theories; they are theories about contingent features of reality. (L) and (CF) are not *causal* explanations; and what they say has nothing do to with anything contingent. They are *conceptual* explanations: they are attempts at conceptual analysis. (CF) doesn't say: 'Given the causal laws that *in fact* prevail, a counter-factual is such and such.' (CF) isn't the kind of thing that gains or loses support with new empirical data. And neither is (L). They are both meant to be non-contingent conceptual analyses. They may be right or wrong. But they are *not* things that are confirmed or disconfirmed by empirical data. So we cannot explain away the circularity in question the way we can explain the circularity (supposedly) characteristic of empirical theories.

And the circularity in question *is* vicious. Part of the reason causal explanations aren't (always) vicious is that causal propositions are contingent; such explanations do not affirm *conceptual* relations among propositions. So the propositions related by a causal explanation do not depend for their *content* on the truth of the causal explanation in question. But conceptual explanation is a different. Where that kind of explanation is a concerned, circularity means a loss of *content* – a loss of intelligibility. The proponent of (L) is trying to give a conceptual explanation of what laws of nature are. The proponent of (CF) is trying to give a conceptual explanation of what counter-factuals are. Somebody who puts forth both (CF) and (L) as conceptual explanations is incapable of even giving *content* to either proposition. (CF) has content only in so far as the concept of (L) has some independent content. And (L) has content only if the concept of a counter-factual has independent content. If each has content only on the condition that the other does, then neither has content. So it is not possible that both (CF) and (L) should constitute conceptual explanations. But that is exactly what

they are supposed to be.

But there is a way out of this for the factualist. He could say:

(SL) I grant that (propositions stating) laws support counter-factuals. But that is not what *makes* something be a law. That is a mere *symptom* of its being a law. It is true that *if* you put a piece of pinewood in water, it will float. But the law *pinewood floats in water* is a different thing from a proposition affirming the existence of that law; laws are not propostions; they are features of the world. In particular, the law *pinewood floats on water* is distinct from the counter-factual proposition: *if you put a piece of pinewood in water, it would float* The just mentioned law is some feature of the world – presumably some kind of glue that connects events. The just mentioned counter-factual is some kind of *consequence* of that law. But is a totally different thing from the law itself. A law is not a proposition at all; it is a feature or constituent of the world. Counter-factuals, on the other hand, are propositions. What *makes* something be a law is that it is a certain kind of feature of the world. Granted, if a thing x is that kind of feature or constituent, then propositions affirming its existence will support counter-factuals. But *this* fact is a mere consequence of the fact that x is a law. It is not what it is for x to *be* a law.

I fully grant that if (SL) is correct, then there is no incoherence or circularity in holding both (CF) and (L). For if (SL) is right, then (L) by no means gives a *conceptual* analysis of what a law is. And it isn't meant to. (L) simply states a true fact about laws. But since it isn't meant to give the *content* of the concept of law. The content of that concept is given by some completely different proposition. So a person could accept *both* (L) and (CF) without being guilty of any kind of vicious circularity. For (L) has content

quite independently of (CF).

The question is not: is (SL) correct? It may well be correct. The question is: Can a *factualist* countenance (SL)?

He cannot. So I will now argue.

If the factualist is to advocate (SL), then he must take the Herculean measure of saying that the concept of natural is to not be analyzed in terms of the concept of a counter-factual. This position may or may not be correct. But it is not an easily defended one. That is why, these days, practically everybody thinks that (L) doesn't merely report of a symptom of natural laws, but gives their essence.

But for the sake of argument, let us suppose (SL) is true. (I myself am completely open on the question whether it is true.) Even so, (SL) is not a position that a factualist could coherently take. (SL) is actually the kind of position that somebody takes when he wants to *get rid of* talk of possible worlds in discussions of counter-factuals.

Consider the counter-factual

(*) If I were to drop this porcelain vase from the top of a tall building, and nothing (other than the side-walk) were to break or otherwise retard its fall, then it would shatter.

This is as true a counter-factual as well find. If (SL) is right, then the everything which makes (*) be true is right here in *this* world. The truth-maker of (*) will have nothing to do with other world at all, and the factualist view is wrong.

Let $L_1 \ldots L_n$ be the physical laws that lead from vase-droppings, like the one described, to vase-shatterings, like the ond just described. Given this, what makes (*) be a true counter-factual? Obviously *one* part of the answer is

that $L_1 \ldots L_n$ are physical laws. There is no doubt that the existence of $L_1 \ldots L_n$ is part of the truth-maker of (*).

Permit me to explain my meaning. Let $P(L_1) \ldots P(L_n)$ be the propositions affirming $L_1 \ldots L_n$. Let P(IC) be the proposition affirming certain initial conditions (a vase is released from the top of a building of a certain height…) And let P(VS) be the proposition *the vase shatters (at time t and place p*) for the appropriate values of p and t.

If *$P(L_1) \ldots P(L_n)$ and P(IC)* did not give high probability to P(VS), that would give conceptually necessary grounds for the *falsity* of (*). Clearly if *$P(L_1) \ldots P(L_n)$ and P(IC) does* entail P(VS), that gives conceptually sufficient conditions for the truth of (*).

So (*) says simply that one true proposition entails another. In general, if the objector is right about laws, then what makes a counter-factual claim true is *not* some occurrence in another world, but is the fact that some puirely logical relation holds between one proposition and another.

Let me sum up this difficult argument. According to one point of view, the concept of law is to be anlaysed in terms fo the concept of a counter-factual: a law is given by a universal statement that supports counter-factuals.

Now *if* this is right, then the factualist analysis is false. The factualist says that counter-factuals register facts about other worlds; and he requires that those other worlds be governed by the *samae laws* as our world. So the factualist analyses law in terms fo counterfactuals, and counter-fatuals in terms of laws. So his analyses are cicular. (We saw earlier why this circularity is not of the innocuous kind that characterizes semantic claims like "bachelors are unmarried males and *vice versa*", and is not of the innocuous kind that, allegedly, characterizes empirical theories.)

The only way to go is for the factualist to say: the concept of *law* is not to be anlaysed in terms fo the concept of a counter-factual. If a proposition

supports counter-fatuals, that is indeed a *symptom* of the fact that it affirms a law. But for a propositoni to affirm the existence of a law is for it to ascribe some kind of constituent or feature to this world; and we can describe that feature or constituent with mentioning counter-factuals. A law is some kind of tie between states of affairs. That tie is something that can be exhaustively described *without* mentioning couinter-factuals -- even though, whenever that tie holds, counter-factuals will also hold.

But if *this* is the right analysis of *law*, then the counter-factuals just become entailment relations hlding between propositios. (More specifically, they become entailments in which the antecedent affirms the existence of laws which hold and the existence of states of affarism which do not hold.) Given this analysis of laws, the meaning of: *if S had been the case, then S* would have been the case* is simply: P(S) and $P(L_1)…P(L_n)$ imply P(S*), where $L_1…L_n$ are the laws that obtain.

There is only one way out for the counter-fatualist: he must find a way to analyse the concept of *law* that (a) does not involve the notin of a counter-factual and (b) does not make law be connections between events in this world. Granted, there is a way to do this: we say that laws are mere regularities of some kind. This was Hume's view. L is a law of nature iff L is some kind of regularity or constant conjunction.

But we've already seen the problem with this view. If laws are mere regularities, then two worlds w and w* can be absolutely alike until time t and yet diverge without limit after t. So if laws are Humean, then the occurrences in other worlds are infinitely more likely to *falsify* our counter-factuals – even ones we know to be true – then to validate them.

Let us sum up. If the fatualist analyses laws in terms of counter-fatuals, then his anlaysis is circular. If he alayses them as mere regularities, then his anlaysis is bankrupt. If he analyses them as ties between events – as things

that *groiund* regularities – then his anlaysis is false.

IX. On our view, there is no circularity at all. For counterfactuals are *just* probability statements, and probability is a notion that can be defined totally independently of the notion of a law; in any case, this is true of probability of the *epistemic* kind. The truth of a probability statement supervenes entirely on what is known to be the case – not on what *would* be the case or what *must* (in a nomic sense) be the case. Suppose you say 'the probability that next ball drawn from the urn is .9'. The truth of that statement derives entirely from what *has* happened and, indeed, from what has been *known* to happen. If it is known that 90% of the balls thus far drawn from the vase have been white, then (ceteris paribus), that probability statement is accurate. Of course, in many cases, knowledge of laws of nature must be taken into account: the statement 'the probability that that bar of steel conducts electricity better than that piece of rubber is 99.9999%' is based on (inter alia) knowledge of laws of nature. But the basis for one's belief in *those laws of nature* ultimately reduces to one's knowledge of what *has* happened. So *ultimately* probability statements are warranted by what *has* happened. Suppose Bob is in world w and Bob* is in world w* and the evidence available to Bob concerning what *has happened* (in w) is identical with that available to Bob* concerning what *has* happened (in w*). In that case, Bob and Bob* cannot diverge in their probability-judgments unless one of them is being irrational (or, in any case, unless one of them is doing a worse job than the other in analyzing the data).[114] So the concept of probability is definable entirely independently of the concept of law and the concept of what *would* have happened (the concept of counterfactuality). So our analysis – which reduces counterfactuals to probability statements – is not guilty of circularity (at least not on those fronts).

I must consider an objection to this:

You've said over and over again that the concept of law is to be alaysed in terms of the concept of a counter-factual. In any case, you've allowed that this idea may be correct. But if it is correct, then your analysis of counter-fatuals is deeply absurd.

According to that analysis, counter-factuals are just probability- or entailment-statements. The counte-factual

(*) *If S had been the case, then S* would have been the case*

(**) means: *P(S), conjoined with* $P_1 \ldots P_n$*, probabilifies P(S*),*

where $P_1 \ldots P_n$ are the propositons one knows to be true, minus the proposition that S was not the case.

So you analyse laws in terms of counter-fatuals. But you anlayse counter-factuals as purely logical relations holding among the proositoins *we know* to be true. So, by implication, you are advocating a kind of idealism or relativism about laws. On your analysis, laws are facts about our *beliefs* about the world, not about the world itself. But obviously laws are facts about the world itself, not about our beliefs.

The objector's point involves a subtle mis-understanding of my position

I agree that laws (or statements affirming them) necessariliy support counter-factuals. I also agree that *if* counter-fatuals make statements about *worlds* – about actual scenarios – *then* those worlds must be governed by the

same laws that govern our world. Consider the counter-factual:

(*) *If Kennedy had not been assassinated, then Kennedy would have been Prsident until ’68.*

It cannot be denied that *if* (*) says something about what happens in some other world w, then w must be governed by the same laws that govern oru world: if the laws in w are sufficiently different, or even minimally different, then Kruschev might have been president in w between ’63 and ’68.

But I myself do *not* think that counter-factuals are to be analyzed in terms of the occurrences in other worlds or scenarios. *A fortiori* I do not hold that the worlds in question need be governed by the same laws. Other worlds, and the laws that govern them, drop out as irrelevant. ON my analysis, (*) is simply a probability statement whose antecendent consists of propositiosn one knows to be true, minus the proposition *Kennedy was assassinated*, and whose consequent is another proositoin. Further, as we saw earlier, the antecedent ultimately doesn’t include any statements of law.

So I *don’t* hold that counter-factuals are to be anlaysed in terms of the concept of law. We thus break the circle which leads from my view to idealism about laws.

At the ame time, my analysis easily accommodates the fact statements affirming laws support counter-factuals. A statement affirming a law at the very least *entails* that some kind of regularity holds. The law *copper conducts electricity* certainly *entails* that t every known piece of copper in fact conducts electricity. (The law is *more* than that regularity. But the law certainly requires the existence of that regularity.) Now the law obviously makes-true the kinds of probability-statements that, in our view, are identical with counter-factuals. On our view, *if x were a piece of copper, then x would*

conduct electrivity means: $P_1 \ldots P_n$ and *x is a piece of copper* entails (or probabilies) *x conduts electricityi*, where $P_1 \ldots P_n$ are the proositons one know to be true minus the proposition *x is not a piece of copper*. Among $P_1 \ldots P_n$ are the proostioins Obviuosly the proposition *every known piece of copper conduts electricity*, conjoined with *x is copper*, gives enormous probability to the proposition *x conducts electricity*. So if we construe counter-factuals as probability- or entailment-statements, then laws *do* support counter-factuals – which is exactly the result we want.

I would like to end this paper by clearing up a confusion. Hume's theory of causation has been heavily critizied. One of the more striking criticisms is this:

Humean laws don't support conter-factuals. Humean laws are just regularities. They are not ties *between* states of affairs. Counter-factuals

VIII. I should address what is, no doubt, the most dissatisfying aspect of our analysis.

At a gut level, it seems undeniable that counterfactuals relate events to events, with the qualification that the events occur in alternative or non-actual worlds. If you say *If Clinton been shot, Gore would have become President* you are saying that one event – Clinton's being shot – causes, or leads to, another – Gore's becoming President. The only difference between that statement and 'ordinary' causal statements is that the former concerns events in other worlds, whereas ordinary causal statements are concerned with events in this world.

There is obviously a lot of truth in this comment. The factualist account is

probably more true to what is going on in people's heads than is my account. But we've seen the problem there. The factualist analysis is totally incoherent. Counterfactuals would (with, at most, a few exceptions) be *logically* absurd if the factualist analysis were right. So in so far as what is going on people's heads is factualist, then people are wrong and incoherent

At the same time, counterfactuals are so endemic to human discourse – and to fruitful, rigorous discourses – that we must explore every possible avenue before declaring that they are all wrong. The only way out, as far as I can see, is to construe them epistemically.

The epistemic account doesn't deal well with that residual feeling that, in making counterfactual claims, we are not dealing with shadowy probabilities but with *states of affairs*, albeit ones that happen somewhere else. But this emotional residue should be taken with a grain of salt. When it comes to modality and determinism, people's intuitions are notoriously prone to error. A lot of this, I think, has to do with the fact that these issues relate to the problem of free will – of the extent to which we are 'free' (whatever that means: a major issue in itself) to do things other than what we will *in fact* do. And, for obvious reasons, we tend to think emotionally, rather than logically, about this problem. In connection with this, we tend to 'metaphysicalize' what are in fact purely epistemic phenomena. Before Kripke came along, people tended to think that water could, in some non-epistemic sense, be something other than H_2O. Kripke[115] showed that this is not the case. It is *epistemically* possible that water should be something other than H_2O.[116] But if water really is H_2O, then there is no circumstance, actual or non-actual, where it is anything other than H_2O. Our initial reaction is to read *metaphysical* openness into what is mere ignorance on our part; we project the epistemic into the outer world. (Really this point goes back to Hume,

though not quite in this form.) If I were to say that this tendency to project is what corresponds to that emotional residue just mentioned – that feeling, not done justice to, by our analysis, that counterfactuals are about *states of affairs*, not pallid probabilities – it would not be completely *ad hoc* on my part.

When people are discussing whether we have 'free will' (whatever that means), someone inevitably says (or thinks) 'I have an unshakeable *feeling* of being free; so I *must* be free.' But one's having that feeling doesn't prove anything at all; that feeling is compatible with one's thoughts and deeds being completely determined – and probably with that person's being un-free on any other construal of that notion. And no philosopher sees any probative value in the proclamation: I *feel* how free I am.[117] Speaking psychoanalytically, it is, I think, worth considering the idea that factualism about counterfactuals is a projection onto the cosmos of that discredited 'feeling of freedom'. Factualism is committed to the idea that things *could* have been different in all but some one isolated respect: the world *could* have been just the same *except* that JFK lived rather than died. We've seen why this is wrong. The purely local changes that factualism needs don't exist. The causal structure of the world is much deeper than that: any given thing is held in place by a network of relations that extend as far in every direction as we can fathom. So factualism needs the causal integrity of the world to be much more superficial and negotiable than it in fact is; it needs the world to have something similar to that openness that *we* think (as people, not as philosophers) underwrites *our* being free. By the same token, if that (putative) openness is to exist in us, it needs the external counter-part of some kind of metaphysical openness: if the causal structure of the world is *too* deep, then *we* are locked in. So it is not hard to see how our desire for freedom is mixed together with a desire for just that superficiality in the

causal order which underpins factualism.

IX. I would like to end by making a point that confirms, though in a minor way, our analysis of counterfactuals. What I'm about to say will have more cache with linguistically oriented philosophers, like myself, than it will with those who are not thus biased.

As Grice and Kripke[118] said, it is always better to avoid supposing that a word is *ambiguous* if there is any other way of accounting for its role in communication. Some epistemologists have suggested that the word 'know' is ambiguous between different meanings. (In one sense of the word 'know', I know that I have two hands. In another sense – the strict Cartesian sense – I don't know this.)[119] As a point about epistemology, this may or may not go through. But, as Kripke points out, it is very hard to believe that word 'know' is literally ambiguous, like the word 'bank'. It is hard to believe that 'know' has different literal meanings in 'I know that my car is in the parking lot' and 'I know that I am conscious'.

The word 'if' seems to have a lot of different meanings in English; and I would assume the same is true of its counter-parts in many other languages. There is the nomic 'if', the 'if' of material implication, the 'if' of formal implication, the epistemic 'if' ('if Joe took the morning flight, then he must be at the hotel by now'), the counterfactual 'if', and probably other kinds of 'if's. In some cases, the 'if' *clearly* expresses a relation between propositions, not between states of affairs: *if* there were a largest prime, then there would be a finite number of primes. Both antecedent and consequent are logically impossible; so the hypothetical doesn't deal with states of affairs, not even possible ones. It can only be dealing with relations between propositions. On the other hand, in some cases, 'if' seems to denote a relation between states

of affairs. If ‘if’ really has a propositional meaning in some cases, and a factualist meaning in others, then it is ambiguous.[120] But surely the word ‘if’ has the same semantic *meaning* in counter-logicals that it has in counter-factuals and counter-nomics. The word ‘if’ is not like the word ‘bank’[121]; it is not as though the word ‘if’ has one meaning in ‘if there is a highest prime, then there are finitely many primes’ and a different meaning in ‘if Bob took his medication more often, he would be more chipper’.

Kripke said that if ‘the’ (in the singular) were really *ambiguous* between different meanings, then there would be languages that one word for the referential ‘the’, and another word for the attributive ‘the’, and other words for any other meanings ‘the’ might have. But there don’t appear to be such languages. So ‘the’ (in the singular) is probably not ambiguous between referential and attributive meanings; it probably has one meaning which pragmatic factors map into different communicated messages. Similarly, I would be *very* surprised (paraphrasing Kripke) if, in Albanian or Maori, there were one word for the epistemic ‘if’ and another for the ‘if’ of formal implication, and so on; my guess is that there is one word which does all these jobs – just like in English. Now since ‘if’ *indisputably* has *an* epistemic meaning (actually, most of its meanings are epistemic); and since, by the argument just given, it probably only has one meaning; it seems reasonable to believe that *all* of its meanings are epistemic. Whenever we have a sentence of the form ‘if…then..’ what is being said is *the truth of P bears positively on the truth of Q* (the bearing is 100% in the case of entailment and up to 100% in cases of probabilification). So even in cases (like counterfactuals) where ‘if’ *seems* to express a relation between states of affairs, as opposed to propositions, it really expresses a relation between propositions. This, of course, is the view we have advocated, though not on such narrowly semantic

grounds.

There is one other point that (mildly) favors the analysis just given. In this paper, we've considered "counterfactual" conditionals (sometimes called "subjunctive" conditionals): "if-then" claims, where the antecedent is known to be false. But there are also "indicative" conditionals, these being "if-then" claims where it is open whether the antecedent is true. Suppose we don't know whether Bob studied or not. In that case, might say: "if Bob studied, he'll get an A. If not, he'll get a C." Here we have an indicative conditional.

I would be quite surprised if the "if" in indicative conditions meant one thing, while the "if" in counterfactual conditionals meant another. Presumably, the same relation holds between antecedent and consequent in both of the following:

(*) "if Bob studied, then he'll get an A"

and

(**) "if Smith hadn't become a drunkard, he would have gone on to become the greatest squash player in the world".

Of course, we *could* say that entirely different relations are being said to hold between antecedent and consequent. But that seems artificial. What is different between (*) and (**) is *not* the relation that holds between antecedent and consequent; the difference is that in the one case, but not the other, it is left open whether the antecedent is true or not.

Now we surely don't want to understand *indicative* conditionals in terms of what obtains in *other* worlds. It is the essence of an indicative conditional to make a statement about *this* world. (*) doesn't say that, in a world w that is

just as much like ours as possible, modulo the fact that in w Bob studied, Bob gets an A in w. (*) says that in *this* world, Bob will get an A *provided* a certain condition is met. (*) says "will", not "would".

So if (*) were about states of affairs – if it said that some state of affairs led to another -- they would be states of affairs that held in *this* world.

Further, we definitely don't want to see (*) as affirming a relation between states of affairs. The truth of (*) doesn't depend on whether Bob actually *did* study. (*) says: *If* Bob studied…not*:* Bob *did* study… If Bob didn't study, that doesn't make (*) false.

Thus (*) cannot be to the effect that some state of affairs leads to another; for the state of affairs that would be associated with the antecedent need not even exist.

It would seem, then, that the right way to analyze (*) is to see it as affirming a relation, not between states of affairs, but between propositions. Presumably, the relation is one of probabilification (the antecedent gives high probability to the consequent).

This is how we analyzed counter-factual conditionals. It seems that, given a conditional, the relation that holds between antecedent and consequent isn't affected by whether that conditional is indicative or subjunctive (counter-factual). Thus we must see counterfactuals as affirming relations of probabilification between propositions. The anti-factualist approach prevails.

References

Blackburn, Simon. 1993. "Morals and Modals." In *Essays in Quasi-Realism.* Oxford: Oxford University Press.

--1984. *Spreading the Word.* Oxford: Clarendon Press.

Blanshard, Bland. 1961. “The Case for Determinism.” In *Determinism and Freedom in the Age of Modern Science.* Sydney Hook (ed.). New York: Collier Books.

van Fraassen, Bas. 1980. *The Scientific Image.* Oxford: Clarendon Press.

Grice, H.P. “Logic and Conversation”. 1990 In *The Philosophy of Language.* A.P. Martinich (ed.). Oxford: Oxford University Press.

Hempel, Carl G. 1966. *Philosophy of Natural Science.* Englewood Cliffs: Prentice-Hall.

Keynes, John Maynard. 1921. London: Macmillan and Co.

Kripke, Saul. 1980. *Naming and Necessity.* Cambridge: Harvard University Press.

--1991“Speaker’s Reference and Semantic Reference”. In *Meaning and Truth: The Essential Readings of Modern Semantics.* Jay Garfield and Murray Kiteley (eds.). Paragon House.

Lewis, Clarence Irving. 1929. *Mind and the World Order.* New York: Dover.

Lewis, David. 1973. *Counterfactuals.* Oxford: Basil Blackwell.

Quine, Willard van Orman. 1982. *Methods of Logic.* Fourth Edition. Cambridge: Harvard University Press.

Sentential Reference

It is clear that at least some sub-sentential expressions refer. What is *not* clear is whether *whole* sentences refer.

Frege famously held that sentences refer to *truth-values.* Others have followed suit. Frege's argument for this view is known as the "Slingshot".

Frege's position is by no means universally accepted. Some hold that sentences do not refer at all. Others, e.g. Barwise and Perry, hold that sentences refer to propositions.

This entire debate needs to re-examined. In this essay, I will show how, given some mainstays of contemporary semantics, a case can be made for the following views:

(*) Sentence-*types* don't refer to anything.

(*) Sentence-*tokens* do not refer to truth-values (though they may refer to other things).[122]

(*) The Slingshot is spurious (for it proves things that aren't true, e.g. your car is identical with Mozart's favorite wig).

The Slingshot, I will argue, involves two major fallacies. One of these was

clearly identified by Barwise and Perry (1983).[123] The other, oddly, has been overlooked completely. Let us begin by discussing why sentence-tokens do not refer to truth-values. Then we will examine the Slingshot.

type-reference versus token-reference

Let us start with some foundational points. The sentence-*type* "I am tired" has a meaning; but its meaning is not a proposition. Its meaning is, very roughly, a function from contexts of utterance to propositions. The meaning of such a *token* is a proposition.[124]

Sentences containing indexicals have a "two-dimensional" structure. There is what indexical sentence-*types* mean (functions that assign propositions to sentence-tokens). And there is what indexical sentence-tokens mean (propositions). So, uncontroversially, *some* sentences have a two-dimensional structure.

What about those sentence-types that lack an indexical-component? Perhaps, it might be suggested, *those* sentence-types have propositions, not functions, for their meanings.

There are two problems with this proposal. First, indexical sentences are the rule, not the exception. All, or almost all, sentences of natural language contain tense-markers and thus indexicals.[125]

Second, even if, in natural language, there *were* non-indexical sentence-types, it would be very strange to suppose that *some* sentence-types had propositions for their meanings, while others had functions from contexts of utterance to propositions for their meanings. Such a supposition would introduce a massive cleavage into a class that, in all likelihood, is unified. In fact, arguably, it would amount to a redefinition of "sentence". There would

be "sentences$_1$" – expressions that have functions from contexts of utterance to propositions for their meanings. This would comprise all, or nearly all, of the natural-language expressions we refer to as "sentences". And there would also be "sentences$_2$": these would be sentence-types that had propositions for their meanings.

Sentences$_1$ and Sentences$_2$ are obviously very different sorts of things. It is hard to believe that the class of "sentences" is so heterogeneous – that "I am tired" is in a different category from "2+2=4".

The natural solution is this. Sentence-types uniformly have functions for their meanings: functions from contexts of utterance to propositions.[126] In some cases, it *seems* as though the sentence-type *per se* has a proposition for its meaning. But that is an illusion.

No one would maintain that the sentence-*type* "I am tired" refers to truth-values or propositions. That type is not associated with any one proposition or, therefore, with any one truth-value. So it doesn't refer to truth-values or to propositions. Given that all natural language sentences are in the same class as "I am tired", it is clear that sentence-*types* don't refer to anything. The question is: what do *tokens* of them refer to? The debate regarding sentence-reference must be reconstructed within this framework.

Once such a two-dimensionalism about sentences is accepted, analogical reasoning leads us say that sentence *tokens* refer to propositions. The type "you" doesn't refer to anything. Its semantic content is a function (from contexts of utterance to individuals). The semantic content of a token of "you" is some individual. The semantic content of such a token coincides with its referent.

The sentential analogue of this point is: The type "I am tired" doesn't refer to anything. Its semantic content is a function from contexts to propositions. The semantic content a token of "I am tired" is the proposition

assigned to it by that function. The semantic content of a *token* of "I am tired" coincides with its referent.

So once we become alive to the two-dimensional nature of sentences, analogical reasoning suggests that sentence-tokens refer to their *meanings* – to the things they are assigned by the corresponding types. Those things are not truth-values. Of course, analogies are not always probative. Fortunately, we can do better than an analogy.

reference and semantic contribution

When we examine the concept of reference, the idea that sentences (or, more precisely, sentence-tokens) refer to truth-values is put under severe strain.

Suppose you want to make a statement about Plato. What you'll do is use some expression that *refers* to him, e.g. "Plato" or perhaps "the author of the Republic". Suppose you want to ascribe wisdom to Plato. You will say "Plato was wise" or "perhaps "the author of the Republic was wise". There is obviously an essential connection between the fact that "Plato" refers to Plato and the fact that, if you want make a statement about Plato, you can use the word "Plato" (in a purely literal manner[127]) to do so.

Suppose that "Plato was wise" meant *Charlemagne was wise.* Suppose that, in general, sentences of the form "…Plato…" meant:…*Charlemagne.*… In that case, it would be very hard to maintain that "Plato" referred to Plato; and it would, to say the least, be natural to suppose that "Plato" referred to Charlemagne.

Suppose you want to make a statement about the truth-value *true.* What you'll do is use some expression like: "the truth-value *true*" or "the property

of being true". Perhaps you'll use some neologism like "**T**". But what you *won't* do is use "snow is white" or "grass is green". Those expressions would be absolutely useless; you couldn't possibly use them to talk about truth-values, except (at most) in some circuitous sense.

Consider the sentence:

(*) "the last thing Plato said had the truth-value *true*."

If "snow is white" really refers to the truth-value true, then we should be able to use that sentence-token to refer to that truth-value. We should thus be able to rephrase (*), using "snow is white" instead of the "truth-value true". But this doesn't seem remotely possible.

If you want to produce a sentence that means: *everything Bob says has the truth-value true*, you'll say "everything Bob says has the truth-value true" or "everything Bob says is true". You *won't* say "everything Bob says has snow is white".

Let us consolidate these points. Sentence-tokens of the form "…snow is white…" don't mean:…*the truth-value* true… "Snow is white and grass is green" doesn't mean: *the truth-value* true *and grass is green.* "Possibly, snow is white" doesn't mean: *possibly, the truth-value* true. By contrast, "Plato is wise" means *Plato is wise.* In general, sentence-tokens of the form "… Plato…" mean:…*Plato.*…This, as we've seen, is indissolubly connected with the fact that "Plato" refers to Plato. So the relationship between a token of "snow is white" and the truth-value *true* isn't remotely like the relationship between "Plato" and Plato.

Of course, there is an obvious reply:

If I say "snow is white", I *am* saying something about the truth-value *true.*

I am saying that *snow is white* has that truth-value. So I *can* use “snow is white” to make a statement about that truth-value. So “snow is white” *does* function like a term that refers to that truth-value.

But this isn’t really responsive to the point we made earlier. For the sake of argument, suppose “snow is white” refers to the truth-value *true*, just like “the truth-value true”. In that case, presumably, “snow is white” ought to be able to perform the functions performed by “the truth-value true”. But this isn’t the case. Consider the sentence:

(lsp) “The last sentence Plato uttered had the truth-value *true*”

If you replace “the truth-value *true*” with “snow is white” or “grass is green”, what results is rubbish:

(lsp*) “The last sentence Plato uttered had snow is white”.

Of course, one can do some syntax-chopping to make the substitution work; one can say that (pt) is equivalent to (say):

(lsp**) “The last sentence Plato uttered had the same truth-value as *snow is white*.”

But clearly the idea that “snow is white” refers to the truth-value *true* is under serious strain. If “snow is white really refers to the truth-value *true*, why is it so hard to it to make it act as though it did? Why is it so hard to make it behave like expressions that we *know*, for certain, to refer to that truth-value?

Of course, one can say:

Semantically ‘snow is white’ is synonymous with ‘the truth-value *true*’. The differences between them that you’ve just described are pragmatic.

But this obviously over-extends the concept of pragmatics. Pragmatics cannot be used to exonerate *every* semantic theory. Suppose I say that “grass is green” means: *penguins don’t fly.* Obviously my theory will conflict with what “grass is green” is usually taken to mean, and that conflict will obviously have a semantic basis.

Suppose I dug in my heels and said:

Semantically my theory is right. Admittedly, the facts appear to be discrepant with my theory. But that has to do with pragmatics, not semantics.

Clearly I would be over-using the concept of pragmatics: the problem is that “penguins don’t fly” and “snow is white” are *too* remote in meaning. “Snow is white” *seems* no less remote in meaning from “the truth-value *true*”. Arguably, they are *more* remote in meaning than the first pair; the first pair consists of two indicative sentences; the second consists of a sentence and sub-sentential expression. So it is hard to see how pragmatics could be used to exonerate the view that those two expressions co-refer.

There is another matter to consider. The semantic content of a token of “you” is the individual assigned to that token by the function encoded in the corresponding type. The semantic content of a token of “I am tired” is the proposition assigned to that token by the function encoded in the corresponding type. In general, the semantic content of an expression-token is the thing assigned to that token by the corresponding type.

The semantic content of a token of “I am tired” is some proposition[128]; it is not the truth-value *true.* The function encoded in the type “I am tired” assigns truth-evaluable entities to the relevant tokens; it assigns things that are true or false. Propositions are true or false. The truth-value *true* is not itself true or false. The semantic content of a (non-abortive) token of “I am tired” or “John is in Cleveland” is something that is true or false. The truth-value *true* is not itself true or false; so it is not the semantic content of any such token.

more reasons to doubt that sentence-tokens refer to truth-values

Consider the sentence-token “George W. Bush smoked a cigarette on April 2, 2005”. You and I both understand that expression as well as it can be understood. We completely and thoroughly grasp its semantics. But we don’t have the slightest idea whether it is true or not.

Except perhaps in very special cases, knowledge of the semantics of a sentence-token absolutely does not require knowledge of its truth-value.[129]

If a term refers, its referent is an integral part of its semantic content, and you cannot possibly know its semantics unless you know what it refers to. Suppose that “Blark” is somebody’s name, but you have no idea whose name it is. Because you don’t know who “Blark” refers to, you really don’t know its semantics. You don’t quite know what is meant by sentence-tokens like “Blark is a nice person”; you don’t know what proposition they encode. Of course, you know that that such sentence-tokens will be true if somebody named “Blark” is a nice person. But the proposition encoded in “Blark is a nice person” is not: *somebody named ‘Blark’ is a nice person.* So if a term refers, one cannot know its semantics without knowing its referent.

Let us apply this point to sentence-tokens. I know the *meaning* – the semantic content -- of a token of "George Bush is currently napping". But I don't know its truth-value. In general, one can know exactly what a sentence-token means – one can know exactly what its semantic content is -- without knowing its truth-value. As we just saw, if a term refers, one cannot possibly have a complete knowledge of its semantic content unless one knows to what it refers. One *can* have a complete knowledge of the semantics of a token of "George Bush is currently napping". If such a token *did* refer to a truth-value one would have to know that truth-value to have a complete grasp knowledge of its semantics. It follows that such a token does *not* refer to a truth-value.

To understand a sentence token, one has to know what *would* make it true, not *whether* it is true. Indeed, language would have little point if one knew the truth-value of every sentence-token one understood. So the truth-value of a sentence-token is not part of its semantic content. It follows, for reasons we've seen, that sentence-tokens don't refer to truth-values.

the objection from definite descriptions

There is an obvious response this line of reasoning:

It is obviously very reasonable to suppose that definite descriptions are singular terms – that "the inventor of bifocals" refers to Benjamin Franklin and "the even prime" refers to the number two. Given this, consider the definite description "the richest man in Holland". You can understand it perfectly without having the foggiest idea to whom it refers. Consider the expressions: "the chemical-compound that cures AIDS", "the sequence of sentences that proves the most important mathematical result of the 23rd century", "the exact number of people in Nepal". You have no idea what

these expressions refer to. But you understand them perfectly. So one *can* grasp the semantics of an expression without knowing its referent. So given only that one can grasp the semantics of a sentence without knowing its truth-value, it does not follow that a sentence's referent is anything other than its truth-value.

This is an important objection. But it is based on a failure (once again) to distinguish type-semantics from token-semantics, and is completely fallacious.

For the sake of argument, let us suppose, with the objector, that definite descriptions are in fact singular terms (as opposed to Russellian quantifiers). Let w be our world and let w* be some other world where English is spoken. So "the inventor of bifocals" has the same *semantics* in both w and w*. Further, suppose that in w* Thomas Jefferson (as opposed to Benjamin Franklin) invented bifocals.

We have a problem. By supposition, "the inventor of bifocals" has the same *semantics* in both w and w*. Presumably, the referent of a referring term is at least part of its semantic content. ("Plato" refers to Plato, and that fact is essential to the semantics of "Plato".) "The inventor of bifocals" refers to one thing in w and to a different thing in w*. Thus, it would appear, "the inventor of bifocals" does *not* have quite the same semantic content in w* that it has in w. So "the inventor of bifocals" both does, and does not, have the same semantics in w and w*.[130] A paradox.

There is an easy way to make sense of all this. Strawson[131] showed us how. Let us suppose, with the objector, that definite descriptions are singular terms. We begin by distinguishing the expression-*type* "the inventor

of bifocals" from *tokens* of that expression. The expression-*type* doesn't refer to anything. But the expression-type does have meaning – it does have semantic content. The meaning of that expression-type is a rule that assigns referents to *tokens* of it, depending on the context of utterance. (In this case, the world of utterance is at least part of the context of utterance.)

So the meaning of the expression-*type* "the inventor of bifocals" is some rule like the following: Given any token of "the inventor of bifocals", if that token occurs in a context where there is exactly one unique bifocal inventor x, then that token refers to x.[132]

Thus, in w, *tokens* of "the inventor of bifocals" refer to one thing (Benjamin Franklin), while in w*, such tokens refer to a different thing (Thomas Jefferson). So the semantic content of a *token* of "the inventor of bifocals" in w is different from the semantic content of a *token* of that expression in w*.

So if, as the objector assumes, definite descriptions are singular terms, there is no one entity that has the same semantics in different worlds while having different referents. The meaning of the expression-*type* "the inventor of bifocals" is the same in both w and w*. But the expression *type* doesn't refer to anything. So, trivially, it doesn't refer to different things in different worlds. *Tokens* of that type are what refer. In w, such tokens refer to Benjamin Franklin. In w*, such tokens refer to Thomas Jefferson.

The principle I am defending is this: If an expression refers to something, it is not possible to have complete knowledge of its semantics without knowing its referent. Definite descriptions are no counter-example to this principle. Definite description-*types* don't refer to anything. So, vacuously, they are no counter-example to our thesis. Tokens of such types do indeed refer. But you cannot possibly know the semantic content of such a token if you don't know what it refers to. The semantic content of a *token* of

“the inventor of bifocals” in w is Franklin. If you are in w and you don’t know what a given token of that expression refers to, then you simply don’t know its semantic content. So, once again, our principle is upheld.

Here is an analogy – I think it is more than an analogy. “You” obviously refers to different people on different occasions. But it doesn’t undergo a change in *meaning* every time it is used to refer to a new person. It is not as though “you” is ambiguous between Fred and Charlie and Mary and Joanne. “You” is not like “bank”.

Here we have two options. One is to say that “you” has a variable reference, but a constant meaning: its referent changes, but not its semantic content.

But we’ve already seen why that would be a tortured and arbitrary position. This brings us to the other option. We say: the *type* “you” always has the same semantic content. That content is a function from contexts of utterance to individuals. A given *token* of you has such an individual for its content.

The type “you” doesn’t refer to anything. So that type is vacuously consistent with our principle that, *if* an expression refers, then one must know to what it refers in order to have a complete knowledge of its semantics.

Tokens of the type “you” do indeed refer. The semantic content of such a token coincides with its referent. So there is plainly no way to know the semantic content of such a token unless one knows to what that token refers. Thus our principle continues to be upheld.

Let us sum up. Whenever we have something that refers, it turns out that one *does* have to know its referent to know its semantics. Apparent examples to the contrary are based on a failure to distinguish type-semantics from token-semantics. One can completely understand – can completely grasp the semantics of -- a token of “George Bush is napping” without knowing its

truth-value. Therefore that token doesn't refer to a truth-value.[133]

do sentence-tokens ever contribute truth-values?

Let us consider an objection to what we've said:

Some connectives, e.g. "and", are truth-functional. "snow is white and grass is green" is true iff both "snow is white" and "grass is green" are true. The truth of the conjunction is a function of the truth-values of its components. So surely the truth-values of "snow is white" and "grass is green" are at least *among* the things which they semantically contribute to "snow is white and grass is green". In general, it seems that, in some contexts, what a sentence semantically contributes is a truth-value. So you're wrong to say that sentences don't semantically contribute their truth-values.

This reasoning is quite spurious. Consider the sentence-token:

(kg) "Kermit is green".

Whether (kg) is true is a function of whether Kermit is green. But the semantic contribution of "Kermit" is not the property of being green: "Kermit" semantically contributes Kermit, not the property of being green. What contributes the property of being green is the expression "green" (or "is green"). So even though the truth of (kg) *depends* on whether Kermit is

green, “Kermit” obviously doesn’t contribute that property to the meaning of (kg).

Now consider the sentence-token:

(js) “John believes that snow is white”.

Whether (js) is true is a function of whether a certain proposition has a certain property: it is a function of whether the proposition *that snow is white* has the property of being believed by John. Plainly “that snow is white” doesn’t semantically contribute the property of being believed by John. What contributes that property is the phrase “is believed by John”.

Now consider the sentence-token:

(sg) “snow is white and grass is green”.

(sg) comprises tokens of “snow is white” and “grass is green”. Whether (sg) is true *depends* on whether the propositions encoded in those sentence-tokens have a certain property (the property of being true), just as the truth of “Kermit is green” depends on whether Kermit has a certain property. But it doesn’t follow that those constituent sentence-tokens *contribute* that property. And, demonstrably, they do not. The sentences themselves no more contribute the property of truth than the occurrence of “Kermit” in (kg) contributes the property of being green and “that snow is white” in (js) contributes the property of being believed by John. In (sg), what contributes the property of truth is the “and”, which can be thought of as a second-level

predicate – as denoting a property of propositions or sequences of propositions.

(sg) is equivalent to

(sg*) and*<snow is white, grass is green>,

where “and*” is a second-level predicate: one that ascribes truth to the propositions encoded in the sentence-tokens falling in its scope.

So the “and” in (sg) can be thought of as ascribing truth to both members of that sequence. [134] So the reason that “snow is white” occurs truth-functionally in (sg) is *not* that it contributes the property of truth; it is that the property of truth is *ascribed* to the proposition it contributes.

These points are worth looking at from a slightly different angle. The truth of (kg) is a function of whether the thing meant by “Kermit” has the property of being green. This is equivalent to saying: (kg) *attributes* the property of being green to the thing meant by “Kermit”. So the semantic role of “Kermit” is *not* to contribute that property; it is to contribute the thing to which that property is being attributed.

The truth of (sg) is a function of whether the things meant by “snow is white” and “grass is green” are true. This is equivalent to saying: (sg) *attributes* truth to those things. So the semantic role of “snow is white” and “grass is green” is *not* to contribute that property. It is to contribute the things to which that property is being ascribed.

(sg) is truth-functional because a certain property is being *ascribed* to the propositions encoded in the constituent sentence-tokens, *not* because those sentences themselves contribute that property. A truth-functional occurrence

of a sentence-token is one in which truth is *attributed* to the corresponding proposition, not one in which that sentence itself contributes the property of truth.[135]

the Slingshot

So far we've seen some reasons to hold that neither sentence-types nor sentence-tokens refer to truth-values. But there is a famous argument to the contrary. It is known as the Slingshot.

I wish to show that the Slingshot is not to be trusted. It proves things that nobody would accept as true, no matter what their semantic views are. It also contains two identifiable errors.

First of all, what is the Slingshot? Let S be any sentence-token, and let P be any proposition. Starting with S, we can produce a sentence-token that has P for its meaning. All we have to do is replace referring terms with co-referring terms. Even if we are limited to that one operation, we can still reach our destination.

Consider the sentence-token "Socrates was wise." Suppose we want to produce a sentence-token that has for at least part of its meaning the proposition: *Mozart wrote fewer fugues than Bach.* And suppose we only allow ourselves to replace referring terms with co-referring terms.

We start with:

(i) "Socrates was wise."

And we replace "Socrates" with a co-referring term:

(ii) "The unique thing x such that Mozart wrote fewer fugues than Bach and x

is Socrates, was wise."

(ii) appears to have, for at least part of its meaning, the proposition: *Mozart wrote fewer fugues than Mozart.*

Replacing a referring term with a *co*-referring term surely doesn't result in a change of *reference*. Let us refer to this principle as (CR).

Given (CR), (i) and (ii) must co-refer. At the same time, (i) and (ii) have different propositions for their meanings. (ii) is about Mozart (among other things), while (i) is not about Mozart.

Whatever it is that a sentence-token refers to, that thing must be invariant with respect to replacements of referring terms, in that sentence-token, with co-referring terms. Propositions are *not* thus invariant. So sentence-tokens don't refer to propositions.

The only thing that is thus invariant is a sentence-token's *truth*-value. Replacing a referring term with a co-referring term never turns truth into a falsity or falsity into truth. So sentence-tokens refer to truth-values, if they refer to anything.

We can make this argument tighter (the tightening is due to Church 1943[136]).

Church 1943[137]

Let S and S* be any two non-analytic sentence-tokens that have the same truth-value.

(1) S
(2) The class of all things x such that (x=x and S) is identical with the class of

all things x such that (x=x).
(3) The class of all things x such that (x=x and S*) is identical with the class of all things x such that (x=x).
(4) S*.

Let us assume that:

(LL) logically equivalent expressions co-refer.

(1) and (2) are logically equivalent. So (1) and (2) co-refer. (3) is what results when a referring term in (2) is replaced with a co-referring term. So (2) and (3) co-refer. (3) and (4) are logically equivalent. So, by (LL), (3) and (4) co-refer. Thus, (1) and (4) co-refer. What we've just said about (1) and (4) can be said (mutatis mutandis) about any two sentences that are like in truth-value. So any two sentence-tokens, alike in truth-value, co-refer.

If we accept only the assumptions just described, we can prove some very interesting things. Davidson[138] "proves" there is only one fact. Pick any two true sentence-tokens. Pick, say, "grass is green" and "whales are mammals". (i)-(iv) are meant to be sentence-tokens, not propositions:

(i) The fact that: grass is green.
(ii) The fact that: the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x).
(iii) The fact that: (x=x and whales are mammals) is identical with the class of all things x such that (x=x).
(iv) the fact that: whales are mammals.

(I) and (II) must co-refer. For given (LL), “grass is green” must refer to the same thing as “the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x)”. So, given (CR), (I) and (II) corefer.

(II) and (III) co-refer for the same reason. And (III) and (IV) co-refer for the same reason. So “the fact that grass is green” refers to the same thing as “the fact that whales are mammals”.

Davidson accepted this conclusion: he thinks there is only one fact. He used this fact (pun intended) as a way of justifying some far-reaching epistemological views.[139]

Here is an analogue of Davidson’s argument: Let (I)-(IV) be sentence-tokens, not propositions.

(I) The proposition: that grass is green.
(II) The proposition that: the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x).
(III) The class of all things x such that (x=x and whales are mammals) is identical with the class of all things x such that (x=x).
(IV) The proposition that: whales are mammals.

(I) and (II) must co-refer. For given (LL), “grass is green” must refer to the same thing as “the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x)”. So, given (CR), (I) and (II) corefer. (II) and (III), and also (III) and (IV), co-refer for the same reason.

So “the proposition that grass is green” refers to the same thing as “the

proposition that whales are mammals". Therefore the proposition that whales are mammals is identical with the proposition that grass is green. Any two true propositions are identical. There is only one true proposition.

But there are many true propositions. Obviously the Slingshot involves a fallacy.

Here is an even more interesting application of the Slingshot. Some philosophers distinguish between constitution and identity. The statue is not *identical* with the clay because you can destroy the statue without destroying the clay. But the clay obviously *constitutes* the statue.

Everyone agrees that every physical entity is either identical with, *or* constituted by, displacements of mass-energy. Your car is *constituted* by a long, complicated series of mass-energy displacements. Let S be a sentence that describes exactly that series. Your favorite shirt is constituted by some entirely distinct series of mass-energy displacements. Let S* be a sentence that describes exactly that series.

For obvious reasons, the series of mass-energy displacements that makes S true is identical with the series that makes S, conjoined with "triangles have three sides", be true.

(A) The set comprising all and only the mass-energy displacements that make true the sentence: S.

(B) The set comprising all and only the mass-energy displacements that make true the sentence: The class of all objects x such that (x is identical with x and S) is identical with the class of all objects x such that (x is identical with x)

(C) The set comprising all and only the mass-energy displacements that make true the sentence: The class of all objects x such that (x is identical with x and S*) is identical with the class of all objects x such that (x is

identical with x)

(D) The set comprising all and only the mass-energy displacements that make true the sentence: S*.

The set comprising all and only the mass-energy displacements that make up your favorite shirt has exactly the same members as the set of such displacements that make up your car. Any two objects are constituted by exactly the same mass-energy displacements. Given any two physical objects, they are either identical, or one of them constitutes the other. In any case, this is what the Slingshot proves.

looking for problems with the Slingshot

The slingshot proves things that aren't true. So we should look for problems with it. If I am not mistaken, there are at least two. The first is identified by Barwise and Perry (1983).

the first problem

Consider the following two sentence-tokens:

(i) Socrates was wise.

(ii) The unique thing x such that Mozart wrote fewer fugues than Bach and x is Socrates, was wise.

There are two ways to read the definite description in (ii). It can be read as a singular term (an expression that refers to an individual), or it can be read as a quantifier.

Referring terms semantically contribute their referents, and nothing else. Given this, suppose we read the definite description in (ii) as a term that refers to an individual. In that case, (ii) has exactly the following proposition for its meaning:

(ii_R) Socrates was wise.

If we treat definite descriptions as expressions whose tokens refer to individuals, then (ii) and (i) have *precisely* the same meaning. Given that a referring term semantically contributes its referent and nothing else, it follows that co-referring terms contribute the exact same thing. Thus, replacing referring terms with co-referring terms cannot change meaning.

Replacing referring terms with referring terms not only preserves truth-value: it preserves *meaning.* Since CR preserves truth-value *and* meaning, we have no grounds for saying that a sentence-token refers to anything *other* than its meaning.

Suppose we treat definite descriptions as quantifiers. Does that make things better for Frege's argument? It does not.

If definite descriptions are quantifiers, then (ii) means:

(ii_{RUS}) For some x, x was uniquely such that [Mozart wrote fewer fugues than Bach and x was Socrates] and x was wise.

Russell thought that quantifiers, and thus definite descriptions, don't refer to

anything. So if Russell is right, then (ii) is not what results when a referring term in (i) is replaced with a co-referring term. If definite descriptions don't refer, then Frege's argument crashes: CR doesn't guarantee the needed steps.

To be thorough, we must consider a nuance. A case can be made that Russell's analysis of definite descriptions actually *is* consistent with the idea that they are referring terms. If he is right, they are, at least arguably, terms that refer to second-level functions.[140] Let C be the class of wise objects. (ii_{RUS}) is true exactly if C contains exactly one object x such that (x is identical with Socrates and Mozart wrote fewer fugues than Bach), and no such object falls outside of C. So if Russell's analysis is correct, the definite description in (ii) can be seen as a function that assigns truth to a class exactly if that class contains a unique x such that (x is identical with Socrates and Mozart wrote fewer fugues than Bach), and no such object falls outside of that class.

If definite descriptions do refer to functions, then (ii) is indeed what results when a referring term in (i) is replaced with another referring term. *But the two terms do not co-refer*. For "Socrates" refers to an individual, and therefore doesn't co-refer with anything that refers to a function.

Let us sum up. If definite descriptions refer to individuals, then (i) and (ii) mean exactly the same thing. If definite descriptions are quantifiers, then (ii) is not what results when a referring term in (i) is replaced with a co-referring term. In either case, Frege's argument is destroyed.

These considerations break Church 1943. (Let Omni be the class of all self-identical objects.) If we read the definite descriptions as expressions that refer to individuals, we have:

(1_R) S
(2_R) Omni=Omni.

(3_R) Omni=Omni.
(4_R) S*.

There is no logical equivalence between the first two steps, or the last two. The argument fails.

So if Church 1943 is to work, the definite descriptions must be read as quantifiers. But then there is no logical equivalence between any two of the entries. Let S be the sentence-token "Mozart wrote fewer fugues than Bach" and let S* be the sentence-token "whales are mammals".

In that case, (2_R) and (3_R) become existence claims that are not logically equivalent. For even though Omni is identical with Omni in all worlds, there are logically possible worlds where whales are mammals and where Mozart wrote more fugues than Bach. (2_R) is contingent on the truth of *Mozart wrote fewer fugues than Bach.* (3_R) is not contingent on that truth. So (2_R) and (3_R) are not logically equivalent.

Let us sum up. There are two ways to read the descriptions in (1)-(4): as singular terms or as quantifiers. If they are read as singular terms, then there is no logical equivalence between the first two entries or between the last two: and the argument fails. If the descriptions are read as quantifiers, then (3) is not what results when a referring term in (2) is replaced with a *co*-referring term: and the argument fails. Basically, as Barwise and Perry observed, once we are clear on how we read definite descriptions, the Slingshot crumbles.[141]

Frege's way out

Frege and Church wouldn't necessarily be moved by the argument just

given; for they wouldn't agree with the conception of reference that underlies it. In this section, I would like to show why their disagreement is innocuous, at least as far as the cogency of our argument is concerned.

Here is the controversial assumption that I made. *If* (e.g.) "Plato" and "the unique x such that 1+1=2 and x=Plato" co-refer, then they both contribute exactly the same thing, namely Plato. Consequently, replacing the one with the other will *not* change sentence-meaning; and this, as we saw, breaks the Slingshot. (If replacing the one with the other *does* change sentence-meaning, that is because they are not *co*-referring terms: one of them is a singular term that refers to a philosopher while the other is a quantifier.)

Let us generalize this. I assumed that if E is an expression that refers to some object O, then when E occurs in a sentence-token, its semantic contribution to the sense of that token is O: nothing more, nothing less. Of course, *if* this is right, *and* E and E* both refer to O, then they both contribute exactly the same thing, namely O.

But Frege (and Church) wouldn't agree with the conception of reference that underlies our attack on his argument. He says that referring terms have both "sense" and reference.[142] A referring term semantically contributes not (only) its referent but its *sense*. Co-referring terms can have different senses. So it *is* possible to change sentence-meaning by replacing referring terms with co-referring terms.

Let us make this more concrete. In Frege's view, "the teacher of Aristotle" and "the author of the Republic" are both referring terms that refers to Plato. So they co-refer. But, according to Frege, they make different semantic contributions. When "the teacher of Aristotle" occurs in a sentence, it contributes some concept to the proposition encoded in that sentence. When "the author of the Republic" occurs in a sentence, it contributes a

different concept.

That is why

(*) “the teacher of Aristotle was wise”

encodes a different proposition from:

(**) “the author of the Republic was wise”.

Let us evaluate this view. Consider the expression “the teacher of Aristotle”. What is the concept contributed by that expression? Here a slight detour may be useful.

Consider the sentences (or sentence-tokens):

(i) “A teacher of Aristotle was wise.”
(ii) “Three teachers of Aristotle were wise.”
(iii) “No teachers of Aristotle were wise.”

Each makes a statement about the concept: *teacher of Aristotle.* (i) says that something instantiating that concept was wise. (ii) says that three things instantiating that concept were wise. (iii) says that nothing instantiating that concept was wise.

Clearly, all of “a teacher of Aristotle”, “no teachers of Aristotle”, and “three teachers of Aristotle” semantically contribute the concept: *teacher of Aristotle*. Perhaps they make other contributions. But that concept is their *common* contribution.

So *if* “the teacher of Aristotle” contributes a concept, it contributes (inter

alia) that concept (*teacher of Aristotle*). Thus, if Frege is right, (*) encodes a proposition that has that concept as one of its constituents; it thus makes a statement about that very concept. In this respect, it is just like (i)-(iii).

Supposing this is right, what is that statement? What does (*) say about the concept: *teacher of Aristotle*?

Russell[143] made it very clear what the answer must be. If the concept *teacher of Aristotle* is uninstantiated, then (*) will be false. On the other hand, if there are multiple (equally contextually salient) instances of that concept, (*) will be false. So that concept must be uniquely instantiated. If that instance should be unwise, then (*) will be false.

Thus, *if* (*) encodes a proposition concerning the concept *teacher of Aristotle*, then (*) is true exactly if that concept is uniquely instantiated and any instance of it is wise.

Even referentialists (those who think that definite descriptions are singular terms) agree that (*) is true exactly under that circumstance.[144] Strawson agrees with Russell that "the king of France is bald" is true just in case there is at least one king of France, at most one of king of France, and any king of France is bald. Russell says that if there is no such object, then the statement is false; Strawson says that if there is no such object, then the statement is abortive. They both agree that it is true exactly if those three conditions are met.

Everybody – referentialists and descriptivists alike -- agrees that *if* "the greatest teacher of Aristotle" contributes a concept or sense, then (*) means:

($*_R$) Exactly one thing was a greatest teacher of Aristotle and any such thing was wise.

But if ($*_R$) gives the meaning of (*), then “the teacher of Aristotle” doesn’t refer to Plato at all. It is a quantifier. So either it doesn’t refer to anything or it refers to a function. If we say that “the teacher of Aristotle” contributes a concept, then we must embrace Russell’s theory. The essence of that theory is: definite descriptions do not refer to individuals; they are quantifiers.

By exactly similar reasoning, if we say “the author of the *Republic*” contributes a sense, we are saying that it is a quantifier. In that case, it either refers to a function or it doesn’t refer to anything.

If “the author of the *Republic*” and “the teacher of Aristotle” contribute different senses, it is because they *don’t* co-refer. If, as Russell thought, they don’t refer to anything, then they don’t *co*-refer. If they refer to functions, they refer to different functions, as we saw earlier. If those descriptions are quantifiers, they are different quantifiers, and thus don’t co-refer.

A corollary is that replacing a definite description with a co-referring definite description cannot result in a change in meaning. Replacing “the author of the *Republic*” with “the teacher of Aristotle” results in a change of meaning only if those expressions are entirely different quantifiers, and thus either denote *different* entities (different functions) or denote nothing at all. So there is no way to change meaning by replacing a definite description with a co-referring definite description.

Admittedly, our argument is not yet airtight. For, conceivably, one could say that “the teacher of Aristotle” semantically contributes *both* Plato *and* the concept *teacher of Aristotle.* But this position is a non-starter. If it is right, then “the teacher of Aristotle is wise” encodes a proposition that concerns that concept *and* that concerns Plato. If “the teacher of Aristotle is wise” encodes a proposition about the concept *teacher of Aristotle*, it means: *exactly one thing was a teacher of Aristotle and any such thing was wise.* If

"the teacher of Aristotle is wise" encodes a proposition about Plato, it means: *Plato was wise.* If it does *both*, then it must mean: *Plato was wise and exactly one thing was a teacher of Aristotle and any such thing was wise.* But that is obviously not what is meant by "the teacher of Aristotle is wise". The very idea that it has that meaning is too tortured to consider.

There is more specific point to make. If definite descriptions were *both* quantifiers *and* terms that refer to individuals, then "the teacher of Plato" and "the author of the Republic" would not, strictly speaking, be co-referring terms. Up to a point, they would co-refer. But past that point, they would either not be referring terms, and thus not *co*-referring terms, or they would refer to different things. In either case, it would be false to say that, on the whole, those descriptions were co-referring terms. Replacing "the teacher of Plato" with "the author of the Republic" would indeed result in a change of meaning, but only because those two expressions would *not* be co-referring terms. Replacing "the inventor of bifocals" with "the first post-master general", or "Plato" with "the unique thing x such that arithmetic is incomplete and x=Plato", would result in a change of meaning, but only because the pairs in questions didn't really co-refer. CR would not permit the substitution of, say, "Plato" with "the unique thing x such that arithmetic is incomplete and x=Plato". So CR would no longer permit exactly the substitutions that the Slingshot depends on.

the second problem with the Slingshot

Some Slingshot-style arguments involve the assumption that logically equivalent *sentences* (or, more exactly, sentence-tokens) co-refer. Church 1943 is an example of such an argument. So is Davidson 1967. Unless it is assumed that logically equivalent sentence-tokens co-refer, these arguments

immediately break down. For those arguments explicitly rely on that assumption to generate the needed steps.[145]

The principle that logically equivalent sentences (sentence-tokens) co-refer is thought to be an innocuous extension of the principle that logically equivalent *sub-sentential* expressions (e.g. “Plato” and “the unique thing x such that 1+1=2 and x-=Plato”) co-refer.

But it is not an innocuous extension of that principle. It is an entirely different principle, and it is almost certainly false.

Let us start with some foundational points. Earlier we argued that sentence-*tokens* are what encode propositions. (Sentence-types encode functions that assign propositions to their tokens.) If this is right, then we shouldn’t say that “sentences” are logically equivalent; we must talk about sentence-*tokens* instead.[146]

Given this, what is it for two sentence-tokens to be logically equivalent? Let T and N be sentence-tokens (not types):

(T) “Triangles have three sides.”

(N) “1+1=2”.

T and N are “logically equivalent”. Each of those tokens semantically contributes a proposition. T and N are logically equivalent in the sense that the one proposition entails the other. For two sentence-tokens to be logically equivalent is for the proposition semantically contributed by the one to entail the proposition semantically contributed by the other.

Now consider the expressions:

(P) “Plato”

and

(P*) “the unique x such that 1+1`=2 and x=Plato”.

P and P* are sub-sentential expressions. They are “logically equivalent”. But what does this mean? It does *not* mean that the things they semantically contribute are logically equivalent. The thing they semantically contribute is Plato. Plato is not a proposition. So he is not logically equivalent with anything, not even himself.

When we say that P and P* are “logically equivalent”, what we mean is: for all values of x, the condition, i.e. the propositional function that a thing must satisfy to be referred to by P entails and is entailed by the condition a thing must satisfy to be picked out by P*. Permit me to elucidate this.

Given the semantic rules of English, the condition that a thing must satisfy to be referred to by P is:

(*) x is identical with Plato.

The corresponding condition for P* is:

(**) x is such that 1+1=2 and x=Plato.

For all values of x, (*) entails and is entailed by (**).

Let us hearken back to T and N. When we say that T and N are “logically equivalent” we are *not* saying that, for all values of x, the propositional function that a thing must satisfy to be referred to by T entails and is entailed

by the condition a thing must satisfy to be picked out by N.

When we say that T and N are “logically equivalent”, we are saying that the propositions they semantically contribute are logically equivalent. We are talking about a relationship that holds among the things semantically contributed by those expressions.

But when we say that “Plato” and “the unique x such that 1+1=2 and x=Plato” are logically equivalent, we are *not* making a statement about the things contributed by those expressions. After all, those expressions contribute an individual; and an individual isn’t logically equivalent with anything.

When we say that “Plato” and “the unique x such that 1+1=2 and x=Plato” are logically equivalent, we are talking about the *conditions* that things must satisfy to be semantically contributed by those expressions; we are not talking about the semantic contributions themselves. But when we say that T and N are logically equivalent, we are talking about the semantic contributions themselves.

So the sense in which T and N are “logically equivalent” is entirely different from the sense in which noun-phrases, like P and P*, are logically equivalent.

Thus, given only that “logically equivalent” *noun* phrases – e.g. “Plato” and “the unique thing x such that 1+1=2 and x is identical with Plato” -- must co-refer, it does not follow that “logically equivalent” sentence-tokens must co-refer. The term “logically equivalent” denotes completely different relations in the two cases.

In fact, it would seem that if two expressions are logically equivalent in the one sense, that practically guarantees that they *won’t* be logically equivalent in the other sense. Logically equivalent sentence-tokens typically do *not* semantically contribute the same thing. A token of “1+1=2”

contributes one proposition; a token of “triangles have three sides” contributes a very different proposition. By contrast, logically equivalent noun-phrases necessarily *do* contribute the very same thing. “Plato” and “the unique x such that 1+1=2 and x=Plato” necessarily *do* contribute the very same thing, namely Plato. Logically equivalent noun-phrases *do* necessarily contribute the same thing. Logically equivalent sentence-tokens typically do *not* do so.

The only sentence-tokens that *are* logically equivalent in the latter sense are those that encode the exact same proposition, i.e. are those that are synonymous, like “snow is white” and “schnee ist weiss”.

Slingshot advocates assume that the kind of “logical equivalence” that obtains between P and P* is identical with the kind that obtains between “1+1=2” and “triangles have three sides”. That assumption is wrong. Without it, many Slingshots fail.[147]

Bibliography

Meaning and Truth. 1991. Jay L. Garfield and Murray Kitely (eds.)
St. Paul, Minnesota: Paragon House

The Philosophy of Language. (1990.) Second Edition A.P. Martinich (ed.).
Oxford: Oxford University Press.

Barwise, Jon and Perry, John.1990/1983. “Semantic Innocence and Uncompromising Situations”. In *The Philosophy of Language.* (1990.) Second Edition A.P. Martinich (ed.).
Oxford: Oxford University Press.

-- *Situations and Attitudes.* 1999. Palo Alto: CSLI Publications.

Blackburn, Simon. 1984. *Spreading the Word.* Oxford: Clarendon.

Carnap, Rudolph. 1947. *Meaning and Necessity.* Chicago: University of Chicago Press.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge: M.I.T. Press.

Church, Alonzo. "Carnap's Introduction to Semantics". *The Philosophical Review.* Vol. LII. No. 3. 1943. 298-304.
--1956. *Introduction to Mathematical Logic.* Princeton: Princeton University Press.

Davidson, Donald. 1984. "True to the Facts". In *Essays on Truth and Interpretation.* Oxford: Oxford University Press.
--1984b/1965. "Theories of Meaning and Learnable Languages".
--1990/1967. "Truth and Meaning". In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.

Donnellan, Keith. 1990/1966. "Reference and Definite Descriptions." In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.

Dummett, Michael 1981. *Frege: Philosophy of Language.* Second Edition Cambridge: Harvard University Press.

Evans, Gareth. 1990/1974. “The Causal Theory of Names”. In A.P. Martinich (ed). *The Philosophy of Language*. Oxford: Oxford University Press.
-- 1982. *The Varieties of Reference.* Edited by John MacDowell. Clarendon: Oxford University Press.

Fodor, Jerry. 1990a. “A Theory of Content I.” In *A Theory of Content and Other Essays.* Oxford: Oxford University Press.
1990b. “A Theory of Content I.” In *A Theory of Content and Other Essays.* Oxford: Oxford University Press.

Frege, Gottlob. 1997/1879. “Begriffsschrift”. In *The Frege Reader.* Edited by Michael Beaney. Oxford: Blackwell.
1997/1891. “On Function and Concept.” In *The Frege Reader.* Edited by Michael Beaney. Oxford: Blackwell.
--1990/1982a. “On Sense and Nominatum.” In A.P. Martinich (ed) *The Philosophy of Language* Oxford: Oxford University Press.
-- 1953/1884. *The Foundations of Arithmetic.* Translated by J.L. Austin. Oxford: Basil Blackwell. Reprinted in 1980 by Northwestern University Press.
-- 1997/1892b. “On Concept and Object.” In *The Frege Reader.* Edited by Michael Beaney. Oxford: Blackwell.

Gödel, Kurt. 1944. “Russell’s Mathematical Logic.” In *The Philosophy of Bertrand Russell.* Edited by Paul Arthur Schilpp. Evanston, Illinois: Northwestern University Press.

Kaplan, David 1964. *Foundations of Intensional Logic.* PhD dissertation. University of California, Los Angeles.

--1990/1970. “Dthat”. In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.
-- 1989. ‘Demonstratives’. In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.
-- 1989b. “Afterthoughts”. In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.

Kripke, Saul. 1999/1971. “Identity and Necessity”. In *Metaphysics: An Anthology.* Edited by Jaegwon Kim and Ernest Sosa. Oxford: Blackwell.
--1972. *Naming and Necessity.* Cambridge: Harvard University Press.
--1991/1977 ‘Speakers Reference and Semantic Reference’. In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.

Lewis, David. 1984. *On the Plurality of Worlds.* Oxford: Blackwell.

Lyons, John. 1977. *Semantics.* Cambridge: Cambridge University Press.

McCawley, James D. 1981. *Everything that Linguists Have Always Wanted to Know about Logic.* Chicago: University of Chicago Press.
--1998. *The Syntactic Phenomena of English.* Chicago: University of Chicago Press.

Mill, John Stuart. 1991. “Of Names and Propositions”. In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.

Montague, Richard. 1974. “The Proper Treatment of Quantification in Ordinary English”. In *Formal Philosophy.* Edited by Richard Thomason. New Haven: Yale University Press.

Neale, Stephen. 1990. *Descriptions.* Cambridge: The M.I.T. Press.
-- 1995. “The Philosophical Significance of Church’s Slingshot.” *Mind*, New Series, Vol. 104, No. 416 (Oct., 1995), 761-825.
--2001. *Facing the Facts.* Oxford: Clarendon Press.

Olson, Ken. 1987. *An Essay on Facts.* Palo Alto: CSLI Publications.

Putnam, Hilary. 1962. "The analytic and the synthetic" In Herbert Feigl and Grover Maxwell, eds., *Minnesota Studies in the Philosophy of Science*, III, Minneapolis: University of Minnesota Press.

Quine, Willard van Orman. 1990/1951. “Two Dogmas of Empiricism.” In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.
--1953. “Reference and Modality.” In *From a Logical Point of View.* Cambridge: Harvard University Press.
--1960. *Word and Object.* Cambridge: The M.I.T. Press.
-- 1966a. “Three Grades of Modal Involvement.” In *The Ways of Paradox.* Cambridge: Harvard University Press.
--1966b. “Vagaries of Definition”. In *The Ways of Paradox.* Cambridge: Harvard University Press.
-- 1966c. “Quantifiers and Propositional Attitudes”. In *The Ways of Paradox.* Cambridge: Harvard University Press.

Russell, Bertrand. 1903. *Principles of Mathematics.* Cambridge: Cambridge University Press.
--1990/1905. "On Denoting". In A.P. Martinich (ed) *The Philosophy of Language* Oxford: Oxford University Press.
-- 1957/1910. "Knowledge by Acquaintance and Knowledge by Description". In *Mysticism and Logic.* Garden City, New Jersay: Doubleday.
-- 1920. *Introduction to Mathematical Philosophy.* London: George Allen Unwin.
--1921. *The Analysis of Mind.* London: George Allen Unwin.
-- 1948. *Human Knowledge: Its Scope and Limits.* London: George Allen Unwin.

Stalnaker, Robert. 1991. "Propositions". In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.

Strawson, Peter. 1990/1950. 'On Referring' in A.P. Martinich (ed) *The Philosophy of Language.* Oxford: Oxford University Press.

Tughendat. Ernst. 1970. "The Meaning of 'Bedeutung' in Frege". In *Analysis*., xxx (1970), pp. 177-189.

Weisler, Stephen. 1991. "An Overview of Montague Semantics." *Meaning and Truth.* 1991. Jay L. Garfield and Murray Kitely (eds.) St. Paul, Minnesota: Paragon House

Whitehead, Alfred North and Russell, Bertrand. (1997/1912). *Principia Mathematica (to *56)*. Cambridge: Cambridge University Press.

Sentence-reference versus Sentence-meaning

Introduction

What does “snow is white” refer to? One plausible answer is:

Nothing – sentences don’t refer. “snow” refers. Perhaps “white” refers. But whole sentences don’t refer.

Frege thought otherwise. He argued that whole sentences refer to things. And he held that those things are *truth*-values.

His argument for this conclusion is known as the “slingshot”.

Slingshot-style arguments have been used to prove many things – some of them very dubious.

What I will argue is this.

The slingshot proves things that aren’t rue. So there is something wrong with it.

The slingshot proves, for example, that there is only one proposition. The slingshot doesn’t just prove that sentences refer to truth-values.

But even the thesis that it refers to truth-values is, I believe, one that conflicts with some principles that practically any semanticists would accept.

So instead of accepting the slingshot, we should look for errors in it. And

it contains at least two identifiable errors. There is no way to fix it so that it doesn't contain those errors.

Thus we are not compelled to accept the deliverances of the slingshot. Given only that the slingshot "proves" that sentences refer to truth-values, it doesn't follow that they actually do so refer.

So what do they refer to?

The short answer is: they refer to propositions.

But the short answer is not quite satisfactory.

type-reference versus token-reference

First of all, we must distinguish between sentence-types and sentence-tokens.

Given some pretty uncontroversial principles – principles that are acceptable to slingshot-friends and foes alike - it is clear that sentence-*types* don't refer to anything. They no more refer than the *type* "you" refers.

The sentence-type "I am tired" has a meaning; but its meaning is not a proposition. Its meaning is a function from contexts of utterance to propositions.

The meaning of such a *token* is a proposition.

Sentences containing indexicals have a "two-dimensional" structure. There is what indexical sentence-*types* mean (functions of some kind). And there is what indexical sentence-tokens mean (propositions).

So, uncontroversially, *almost* all sentences have a two-dimensional structure.

What about those sentence-types that lack an indexical-component? Perhaps *those* sentence-types have propositions, not functions, for their meanings?

There are two problems.

First, indexical sentences are the rule, not the exception. All, or almost all, sentences of natural language contain tense-markers and thus indexicals. [148]

(When philosophers wish to discuss genuinely "eternal" sentences, they always have to resort to *ad hoc* stipulations; they say "we are taking the 'is' in a timeless sense"; in effect, they are not considering sentences of natural language, but near cousins.)

Second, even if, in natural language, there *were* non-indexical sentence-types, it would be very strange to suppose that *some* sentence-types had propositions for their meanings, while others had functions for their meanings.

Such a supposition would introduce a massive cleavage into a class that, in all likelihood, is quite unified. In act, arguably, it would amount to a redefinition of "sentence". There would be "sentences$_1$" – expressions that have functions from contexts of utterance to propositions for their meanings. (This would comprise all, or nearly all, of the natural-language expressions we refer to as "sentences".)

And there would also be "sentneces$_2$": these would be sentence-types that had propositions for their meanings.

Sentences$_1$ and Sentences$_2$ are obviously very sorts of things. It is hard to believe that the class of "sentences" is so heterogeneous – that "I am tried" is in an different category from "2+2=4".

The natural solution is this. Sentence-types uniformly have functions for their meanings: functions from contexts of utterance to propositions. In some cases, it *seems* as though the sentence-type *per se* has a proposition for its meaning. But that is an illusion.

An analogy might help. Where addition is concerned, we can drop the

parenthesis; the commutative law guarantees that (a+b)+c and a+(b+c) are the same. But the operations are as distinct as (a÷b) ÷c and a÷(b÷c).[149]

Similarly, there is no less a difference in meaning between the *type* "two and two equals four" and its tokens than there is between the type "I am tired" and *its* tokens.

No one would maintain that the sentence-*type* "I am tired" refers to anything. That type is not associated with any one proposition or, therefore, with anyone truth-value. It obviously doesn't refer to truth-values or to propositions.

Thus, given that all natural language sentences are in the same class as "I am tired", it is clear that sentence-*types* don't refer to anything. The question is: what do *tokens* of them refer to?

The debate the between the slingshot-advocate and the slingshot-foe must be reconstructed within this framework.

Once such a two-dimensionalism about sentences is accepted, analogical reasoning leads us say that sentence *tokens* refer to propositions. The type "you" refers to a function. Tokens of it refer to people. The type "the dog in the corner" refers to functions. Tokens of it refer to dogs. The type "I am tired" refers to functions; tokens of it refer to propositions.

But, I grant, the analogy is less than probative.

Let us see what the Slingshot has to say.

the slingshot

First of all, what is the slingshot?

Let S be any sentence/sentence-token, and let P be any proposition.

Starting with S, we can produce a sentence has P for its meaning. All we

have to do is replace referring terms with co-referring terms. Even if we are limited to that one operation, we can still reach our destination.

Consider the sentence/sentence-token "Socrates was wise." Suppose we want to produce a sentence that has for at least part of its meaning the proposition: Mozart wrote fewer fugues than Bach. And suppose we only allow ourselves to replace referring terms with co-referring terms.

We start with:

(i) Socrates was wise.

And we replace "Socrates" with a co-referring term:

(ii) The unique thing x such that Mozart wrote fewer fugues than Bach and x is Socrates, was wise.

(ii) appears to have, for at least part of its meaning, the proposition: Mozart wrote fewer fugues than Mozart.

Replacing a referring term with a *co*-referring term surely doesn't result in a change of *reference*. Let us refer to this principle as (CR).

Given (CR), (i) and (ii) must co-refer.

But they have different propositions for their meanings. (ii) is about Mozart (among other things), while (i) is not about Mozart.

So sentences/sentence-tokens don't refer to propositions. Whatever it is that sentences/sentence-tokens refer to, that thing must be invariant with respect to replacements of referring terms, in that senence-token, with co-referring terms.

The only thing that is thus invariant is a sentence's (sentence-token's) *truth*-value. Replacing a referring term with a co-referring term never turns

truth into a falsity or falsity into truth.

So sentence-tokens refer to truth-values, if they refer to anything.

We can make this argument tighter (the tightening is due to Church 1943: I am innocuously modifying his argument))

Church 1943[150]

Let S and S* be any two non-analytic sentences/sentence-tokens that have the same truth-value.

(1) S
(2) The class of all things x such that (x=x and S) is identical with the class of all things x such that (x=x).
(3) The class of all things x such that (x=x and S*) is identical with the class of all things x such that (x=x).
(4) S*.

Let us assume – what surely cannot be denied – that

(LL) logically equivalent expressions co-refer.

(1) and (2) are logically equivalent. So (1) and (2) co-refer.

(3) is what results when a referring term in (2) is replaced with a co-referring term. So (2) and (3) co-refer.

(3) and (4) are logically equivalent. So they co-refer.

So (1) and (4) co-refer. Any two sentences, alike in truth-value, co-refer.

If we accept only the assumptions just described, we can prove some

very interesting things. Davidson[151] "proves" there is only one fact. Pick any two true sentences/sentence-tokens. Pick, say, "grass is green" and "whales are mammals". (i)-(iv) are meant to be sentences (or sentence-tokens), not propositions:

(i) The fact that: grass is green.
(ii) The fact that: the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x).
(iii) The fact that: (x=x and whales are mammals) is identical with the class of all things x such that (x=x).
(iv) the fact that: whales are mammals.

(I) and (II) must co-refer. For given (LL), 'grass is green" must refer to the same thing as "the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x)". So, given (CR), (I) and (II) corefer.

(II) and (III) co-refer for the same reason. And (III) and (IV) co-refer for the same reason.

So "the fact that grass I green" refers to the same thing as "the fact that whales are mammals".

Davidson accepted this conclusion: he thinks there is only fact (or no facts). He used this "fact" (pun intended) as a way of justifying some counter-intuitive epistemological views.

Here is an analogue of Davidson's argument: Let (I)-(IV) be sentences/sentence-tokens, not propositions.

(I) The proposition: that grass is green.

(II) The proposition that: the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x).
(III) The class of all things x such that (x=x and whales are mammals) is identical with the class of all things x such that (x=x).
(IV) The proposition that: whales are mammals.

(I) and (II) must co-refer. For given (LL), ‘grass is green” must refer to the same thing as “the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x)”. So, given (CR), (I) and (II) corefer.

(II) and (III) co-refer for the same reason. And (III) and (IV) co-refer for the same reason.

So “the proposition that grass I green” refers to the same thing as “the proposition that whales are mammals”. So the prpostion that whales are mammals is identical with the proposition that grass is green. Any two true propositions are identical. There is only one true proositions.

But there are many true prooistions. So something must be wrong with the slingshot.

There are two problems with the Slingshot. The first is identified by Barwise and Perry (1990/1983).

The first problem

(i) Socrates was wise.
(ii) The unique thing x such that Mozart wrote fewer fugues than Bach and x is Socrates, was wise.

There are two ways to read the definite description in (ii). It can be read as a term that refers to individuals, or it can be read as a quantifier.

Referring terms contribute their referents, and nothing else. So read as a term that refers to individuals (or, to be precise, as a term whose tokens refer to individuals), (ii) has exactly the following proposition for its meaning:

(ii_R) Socrates was wise.

So, if we treat definite descriptions as expressions that refer to (or whose tokens refer to) individuals, then (ii) and (i) have *precisely* the same meaning. And there doesn't seem to be any possibility of change in sentence-meaning if we confine ourselves to replacing referring terms with co-referring terms.

If we treat definite descriptions as quantifiers, things are no better.

If Russell is right, then (ii) means:

(ii_{RUS}) For some x, x was uniquely such that [Mozart wrote fewer fugues than Bach and x was Socrates] and x was wise.

Russell thought that quantifiers (and definite descriptions) don't refer to anything. So if Russell is right, then (ii) is not what results when a referring term in (i) is replaced with a co-referring term.

These days, many think that quantifiers (and some include definite descriptions among them) refer to functions. If that is right, then (ii) is not what results when a referring term in (i) is replaced with a co-referring term. For "Socrates" refers to an individual, and doesn't co-refer with anything that refers to a function.

These considerations break Church 1943. (Let Omni be the class of all self-identical objects.) If we read the definite descriptions as expressions that refer to individuals, we have:

(1_R) S
(2_R) Omni=Omni.
(3_R) Omni=Omni.
(4_R) S*.

There is no logical equivalence between the first two steps, or the last two. The argument fails.

So if Church 1943 is to work, the definite descriptions must be read as quantifiers.

But then there is no logical equivalence between any two of the entries. Let S be the sentence "Mozart wrote fewer fugues than Bach" and let S* be the sentence "whales are mammals".

In that case, (2_R) and (3_R) become existence claims that are not logically equivalent. For even though Omni is identical with Omni in all worlds, there are logically possible worlds where whales are mammals and where Mozart wrote more fugues than Bach. (2_R) is contingent on the truth of *Mozart wrote fewer fugues than Bach.* (3_R) is contingent on the truth of *whales are mammals.* There are worlds where whales are mammals, but where Bach didn't write any fugues. So (2_R) and (3_R) are not logically equivalent.

Basically, once we are clear on how we read definite descriptions, the slingshot crumbles. [152]

the second problem with the Slingshot

Some slingshot-style arguments (though perhaps not all) involve the assumption that logically equivalent *sentences* co-refer.

Church 1943 is an example of such an argument. Unless it is assumed that logically equivalent sentences co-refer, the argument immediately breaks down. The same is true with Davidson 1967.

The principle that logically equivalent *sentences* co-refer is thought to be an innocuous extension of the principle that logically equivalent *sub-sentential* expressions co-refer.

It is not an innocuous extension of that principle. And it is false.

Consider the expressions

(P) “Plato”

and

(P*) “the unique x such that 1+1`=2 and x=Plato”.

P and P* are “logically equivalent”. But what does this mean? It does *not* mean that the things they semantically contribute are logically equivalent. The thing they semantically contribute is Plato. Plato is not a proposition. So he is not logically equivalent with anything, not even himself.

When we say that P and P* are “logically equivalent”, what we mean is: for all values of x, the condition (the propositional function) a thing must satisfy to be picked out by P entails and is entailed by the condition a thing must satisfy to be picked out by P*. Given the semantic rules of English, the condition a thing must satisfy to be picked out by P is:

(*) x is identical with Plato.

The corresponding condition for P* is:

(**) x is such that 1+1=2 and x=Plato.

For all values of x, (*) entails and is entailed by (**).

But the point is: the things that P and P* semantically *contribute* are not logically equivalent.

However, when we say that sentences (or, if we are precise, sentence-tokens) are "logically equivalent", we *are* saying that the things they semantically contribute entail each other. And we are *not* saying that, for all values of x, the condition a thing must satisfy to be picked out by the one expression entails and is entailed by the corresponding condition for the other.

"Triangles have threes" sides is "logically equivalent" with "1+1=2". (More exactly, any token of the one is logically equivalent with any token of the other.) What does the term "logically equivalent" mean here? It means that the thing semantically contributed by the one sentence entails, and is entailed by, the thing semantically contributed by the other. "Triangles have threes" means, i.e. semantically contribute, *triangles have three sides.* "1+1=2" means, i.e. semantically contributes, *1+1=2.* The one proposition entails, and is entailed by, the other.

Given only that "logically equivalent" *noun* phrases – e.g. "Plato" and "the unique thing such that 1+1=2 and x=Plato" -- must co-refer, it does not follow (and it is not the case: see below) that "logically equivalent" sentences must co-refer. The term "logically equivalent" denotes completely different

relations in the two cases.

Slingshot advocates assume that the kind of “logical equivalence” that obtains between P and P* is identical with the kind that obtains between “1+1=2” and “triangles have three sides”. That assumption is wrong. And without it, many slingshots – Church 1943, Gödel 1944, Quine 1953, Davidson1967 – fail.

the positive account

So what *do* sentences/sentence-*tokens* refer to?

The right answer is given to us through a point made earlier. The meaning of the *type* “you” is a function, not an individual. The semantic content of a *token* of “you” is an individual. To understand an utterance is to knows its semantic content. You obviously cannot know the semantic content of a *token* of an indexical, like “you”, unless you know what it refers to. Indexical-*tokens* are concerned, semantic content equals referent.

If definite descriptions are non-quantifiers, then the same thing is true of them. (Actually the same thing is true even if they are quantifiers. But let us leave that aside for the moment.) In that case, the semantic content of the *type* “the inventor of bifocals” is some function from contexts (worlds) to individuals. In any world where English is spoken, the *type* “the inventor of bifocals” has the same semantics it has here: its semantic content is that function.

Obviously in some of those worlds, somebody other than Benjamin Franklin invented bifocals. We don’t want to say that the *type* “the inventor of bifocals” can keep the same semantics while having a variable reference. Given any expression that *has* a referent, its referent – its semantic contribution -- is, one would think, an integral part of its semantics. If that is

right, then it is self-contradictory to say that the very same expression can refer to different things in different worlds and yet have the very same semantics in all those worlds.

One can take the heroic measure of saying that the referent of an expression is not a part of its semantic content. But that is not an easy case to make. Given what semantic content is, it is (in my view) a self-contradictory position.

Fortunately, we don't have to take that heroic measure. We say, simply, the *type* "the inventor of bifocals" has the same semantics in every world where English is spoken. But in some of those worlds, *tokens* of that type refer to people other than Benjamin Franklin.

If this is right, then one can no more know the semantic content of a *token* of a definite description without knowing its referent than one can know the semantic content of a *token* of an indexical without knowing *its* referent. One can understand such tokens only in the pickwickian sense that one understands the corresponding types.

The moral is that, if an expression refers to something, one cannot understand it without knowing its referent. When one *seems* to understand an expression – e.g. "the inventor of bifocals", "the first thing Plato said to Socrates" – without knowing its referent, what is going on is that one understand the expression-*type*; one does not understand the expression-token. And the expression-*type* has no referent; only tokens do.

If sentence-tokens refer to things, one cannot understand them without knowing what their referents are, anymore than one can understand a token of "you" without knowing what *its* referent is.

Obviously we can understand sentence-tokens without knowing their truth-values. So sentence-tokens don't refer to truth-values.

The referent of a sentence-token must be something that one knows

whenever one understands a sentence.

To understand a sentence token, one must know what *proposition* it encodes; one needn't know its truth-value. Indeed, language would have little point if one knew the truth-value of every sentence-token one understood.

So that is one consideration in favor of the view that sentence-tokens do not refer to truth-values.

Here is another. The referent of a referring term is, presumably, the thing it contributes to the meanings of sentences/sentence-tokens in which it occurs. "Plato" refers to Plato because, and only because, when "Plato" occurs in a sentence-token, it makes that sentence –token be about Plato. If occurrences of "Plato" made the corresponding sentence-tokens be about Socrates, then "Plato" would refer to Socrates.

For "Plato" to refer to Plato is for sentence-tokens of the form:

"…Plato…" to mean:….*Plato.*…

In general, for "E" to refer to O is for sentences of the form "…E…" to mean:…*O.*…

Sentences of the form "…snow is white…" mean…*snow is white.*…

"Possibly, snow is white" means: *possibly, snow is white.* "Bob thinks that snows is white" means: *Bob thinks that snow is white*. And "snow is white" – a limiting case of a sentence having the form "…snow is white…" – means: *snow is white.*

Sentences of the form "…snow is white…" don't mean:…*the truth-value* true…

If you say: "possibly snow is white", you are not saying: *possibly, the truth-value* true.

Sentences contribute the propositions they mean, not the truth-values of

those propositions.

Reference *is* semantic contribution. For "Plato" to refer to Plato just *is* for sentences/sentence-tokens of the form "…Plato…" to mean:….*Plato*.…

Sentences of the form "…snow is white…" mean:…*snow is white*, no less than "…Plato…" means:…*Plato*.…

So sentences/sentence-tokens refer to propositions.

It is said that, in some contexts, "snow is white" contributes its truth-value. I think this is a mistake.

Consider "snow is white and grass is green". This is logically, and (I will argue) semantically equivalent to: "the sequence <*that snow is white, that grass is green*> is such that both its members are true."

So *if* "snow is white and grass is green" is to be *true*, then its constituent propositions must be true. (That is why we say that "and" is truth-functional.) But this doesn't mean that the constituent sentences contribute *truth*-values. If "Kermit is green" is to be *true*, then Kermit must be green. But it doesn't follow, and it isn't the case, that "Kermit" contributes the property of greenness. Obviously if

(#) "the sequence <that snow is white, that grass is green> is such that both its members are true."

is to be true, then "that snow is white" and "that grass is green" must both be true. But, just as obviously, that doesn't mean that they both semantically *contribute* the truth-value *true*. If, as I maintain, (#) is equivalent with

(##) "snow is white and grass is green"

then the constituent sentences (or sentence-tokens, rather) in (##) don't

contribute truth-values; they contribute propositions.

If, as I will argue, reference reduces to semantic contribution, then “snow is white” refers to the proposition: *snow is white.*

Carnap (1947) said that the *extension* of a sentence is its truth-value, and that its *intension* is the proposition it means.

This position implodes the second we take a two-dimensionalist view of sentences. For reasons just stated, the meaning of any indexical sentence-type – and this means *any* sentence-type -- is a function from contexts to propositions. So the *intension* of a sentence-type is obviously a function, not a proposition.

Sentence-types don’t have *extensions*. Only expression-tokens have extensions. Let w* be a world where English is spoken, and where there are more whales than there are here. The extension of “whale” in w* is different from the extension of that expression here. But its semantics is the same.

Obviously different *tokens* of the word “whale” have different extensions. The word “whale” does not *itself* have an extension. AT most, it has different extensions in different contexts (different worlds, different times). But this is just an imprecise way of saying that different *tokens* of it have different extensions. The *type* has an intension: a function from contexts (worlds) of utterance to classes. Tokens of it have extensions. The extension of a particular token of is the thing that the function associated with the corresponding type associates with that token.

By parity of reasoning, it is not sentence-*types,* but sentence-tokens, that have extensions. And the extension of a sentence-token is the thing that the function associated with the corresponding type assigns to it. That thing is a proposition, not a truth-value.

To sum up, sentences refer to propositions. If sentences have extensions, their extensions are propositions. (I leave it open whether “extension” and

"referent" are synonyms.)

There is a wrinkle. The expression "that snow is white" *uncontroversially* refers to the proposition *that snow is white*. That expression is neither true nor false. By contrast, "snow is white" is true or false.

It seems strange that, of two co-referring expressions, one should be true or false, while the other should be neither. This is a problem for my analysis.

I will argue that, while there is a semantic difference, between (tokens of) "snow is white" and "that snow is white", it cannot be a difference in what they *refer* to. The difference is a difference in force, not in referent, not in the proposition contributed. And, as I will try to show, differences in force cannot amount to differences in referent.

A sentence, I will argue, is an expression that has force that refers to a proposition. "Snow is white" is what results when "that snow is white" falls within the scope of a force-operator. (In this case, the operator gives *assertoric* force.) "Is snow white" is what results when "that snow is white" is pre-fixed by a different force-operator.

Of course, at the level of surface structure, the expression "that snow is white" is absent from both "snow is white" and "is snow white?". But that means very little. "Expression" is a semantic term; expressions are individuated by their meanings, not their phonetics. Phonetically, my pronunciation of "Socrates" will be very different from a Scotsman's pronunciation; and neither will have *any* appreciable similarity to, say, Stephen Hawking's pronunciation. What makes all of these utterances be tokens of the same word is not phonetics alone. It is phonetics plus various psychological, cultural, and historical factors. The phonetics are really just a way of indicating that these *other* factors are at work. Similarly, what makes an utterance comprise an occurrence of "that snow is white" is not phonetics alone; it is phonetics plus other factors. Given an adequate conception of

what an expression is, our analysis proves tenable, despite its being mildly discrepant with facts about phonetics.

Part I. dismantling the slingshot

Chapter 1. Introduction

"Socrates" stands for Socrates. "Plato" stands for Plato. It is obvious that *sub*-sentential expressions may stand for things.

But do *whole sentences* stand for things? What does "Caesar died" stand for? There are a few obvious answers to this question.

(i) It stands for the *fact* that Caesar died.

(ii) It stands for the death of Caesar.

(iii) It stands for the proposition: *Caesar died.*

(iv) It doesn't stand for anything: reference is a *sub*-sentential notion; it makes no sense to say that whole sentences refer to things.

I think that (i) and (ii) are the same answer. But I won't press that.

Frege[153] argued that "Caesar died" stands for the truth-value *true* (the property of being true). Frege argued that all true sentences stand for the same thing – the truth-value *true* – and that all false sentences stand for the same thing – the truth-value *false.*

Frege's argument for this thesis is as famous as the thesis itself.

Church[154] improved this argument of Frege's. Church's improved version of Frege's argument is known as the "slingshot". Like Frege, Church took sentences to stand for truth-values.

The slingshot has been used to prove a number of astounding things. Here is a sample:

(*) There is only one fact.[155]
(*) There is no distinction between a proposition's *actually* being true and its *necessarily* being true.[156]
(*) The sentence "1+1=2" has one meaning when it occurs on its own, and a completely different meaning in "necessarily, 1+1=2".[157]

The slingshot can also be used to prove (but has *not* been used to prove) some even more astonishing things:

(*) there is only one true proposition.
(*) there is only one truth.
(*) all true sentences have the same meaning.
(*) Anyone who knows anything knows everything.

If the slingshot proves anything at all, then it proves that there is only true proposition. (Thus, since I know my own phone number, I know everything.) *If* we concede that the slingshot shows anything, we must concede that there is only one truth.

Here are some points I will argue for in this work.

(*) Sentences do not refer to truth-values.
(*) Sentences *do* refer to propositions.
(*) The slingshot involves some dubious semantic views, and its probative value is open to serious doubt.

Intuitively, *whole* sentences seem to be in a different category from expressions like “Plato” and “Caesar”. Clearly “Caesar died” *expresses* a proposition; it has a proposition for its meaning or sense. But it seems unnatural to suppose that “Caesar died” refers to something in the sense in which “Plato” refers to something. Presumably the concept of *representation* applies to both cases. But we seem to be dealing with quite different *types* of representation.

I will argue that “Caesar died” refers to the proposition *Caesar died* in *exactly* the sense in which “Caesar” refers to Caesar.

Consider an expression that uncontroversially refers to the proposition: *Caesar died*. Here is an example: “that Caesar died”.

I will argue that “Caesar died” and “that Caesar died” both refer to the death of Caesar in the exact same sense. It isn’t that in some *extended* sense of the term “refer”, “Caesar died” refers to the proposition *Caesar died.* No -- that sentence refers to that proposition in the very sense in which “Caesar” refers to Caesar.

There is an obvious response:

Consider the expression: “Caesar died”. That expression is *true or false.* Consider the expressions “the death of Caesar”, “dthat [the death of Caesar]”, and so on. Those expressions are not true or false. In general, referring terms

are not true or false. “Socrates” is not true or false.

Since referring terms are not true or false, it follows that sentences are not referring terms. (For sentences are true or false.)

This is an important point. We will deal with it at length.

Chapter 2. the slingshot: some misgivings

Frege’s argument[158] (the “slingshot”) seems to involve only two, seemingly incontrovertible principles.

(CR) Replacing expressions with co-referring expressions cannot result in a change of *reference.*

(LL) Logically equivalent expressions cannot differ in reference.

In a moment, we will see that, in fact, a *third* principle is also needed, namely:

(SS) The *sense* of a sentence can be changed by substituting the referring terms in that sentence with co-referring terms.

Typically (SS) is not explicitly cited as a premise in slingshot-style arguments. But, I will argue, it is always relied on.

Some examples may help illustrate what these principles amount to. Consider the expression:

(a) “the wife of the father of the greatest hemlock-drinking philosopher of antiquity.

Suppose we replace “the greatest hemlock-drinking philosopher of antiquity” with a co-referring expression, say “the philosopher who figures as the protagonist in most of Plato’s dialogues”.

(b) “the wife of the father of the great Greek philosopher who figures as the protagonist in most of Plato’s dialogues”.

(*) and (**) obviously have different *senses.* But they *refer* to the same person. In any case, they refer to the same person *if* definite descriptions refer to individuals. And slingshot arguments always involve the assumption that definite descriptions do so refer.

Replacing referring terms with co-referring terms cannot change reference. Let E be a referring expression that *contains* a referring expression e. Let e* be an expression that refers to the same thing as e. Suppose that, in E, we replace e with e*. Let E* be the result. E* must have the same referent as E.

This, I think, should make it clear what (CR) amounts to.

(LL) is pretty transparent. For the sake of argument, suppose that definite descriptions are referring terms. In that case “Plato” and “the unique x such that x=Plato” and “the unique x such that 1+1=2 and x=Plato” must all refer to the same thing, the reason being that they are “logically equivalent”.

Caution is required here. Logical equivalence is typically thought of as a relation that holds between propositions or open-propositions. (By an “open-

proposition" I mean what Russell called a "propositional function", e.g. *x is a man*.) So *all ravens are black* and *all non-black things are non-ravens* are logically equivalent. And so are *x is black if a raven* and *x is a non-raven if non-black.*

Given that logical equivalence is, typically, a relation between propositions or sentences, in what sense can referring expressions be logically equivalent? There is a clear sense in which sentences can be logically equivalent: they can encode logically equivalent propositions. But how can *referring* terms be logically equivalent?

The answer seems to be this. Two referring expressions are logically equivalent iff the condition (the open proposition) that a thing must satisfy if the one expression is to pick it out is logically equivalent to the condition that a thing must satisfy if the other expression is to pick it out.

Let "ze phi" be an expression that *refers* to an object iff that object is the unique phi. (So "ze phi" has the meaning that Strawson[159] attributes to "the phi".) In that case "Plato" and "ze unique thing x such that x=Plato and arithmetic is incomplete" are logically equivalent. To be picked out by "Plato" a thing must satisfy the open-proposition *x=Plato.* To be picked out by "ze unique thing x such that x=Plato and arithmetic is incomplete" a thing must satisfy the open-proposition *x is identical with a unique thing y such that y=Plato and arithmetic is incomplete.* And those two open propositions are logically equivalent.

So (LL) really amounts to this: two expressions must co-refer iff the condition (the open proposition) that a thing must satisfy to be picked out by the one is logically equivalent to the open proposition that a thing must satisfy to be picked out by the other.

Thus stated, it is very hard to see how (LL) could be false.

(SS) requires little elucidation. Consider the sentences:

(c) “the wife of the father of the greatest hemlock-drinking philosopher of antiquity thought her son had bad values.”

Suppose we replace “the greatest hemlock-drinking philosopher of antiquity” with the great Greek philosopher whose dialogues, though estimable, bear no resemblance to actual conversation”.

What results is:

(d) “the wife of the father of the great Greek philosopher who figures as the protagonist in most of Plato’s dialogues”.

Let us now put forth a (broadly) Fregean argument for the position that sentences refer to truth-values. The argument involves two steps. The first step involves showing that sentences *do not* refer to propositions; so the first step establishes a purely negative result. The second step uses the first step as a means to the positive conclusion that sentences *do* refer to truth-values.

The first step in Frege’s argument

Consider sentences (c) and (d). By (CR), they cannot differ in their *referent.*

Now the fact that (c) and (d) have the same referent seems to show that the referent of a sentence is *not* the proposition it encodes. For it seems quite plain that (c) encodes a different proposition from (d). (c) encodes a proposition about (inter alia) hemlock-drinking; (d) encodes a proposition about somebody’s inability to write authentic-sounding dialogue. So given

that (c) and (d) must have the same referent, and given *also* that they encode different propositions, it *does* follow that what a sentence refers to is not the proposition it bears.

Frege's need for (SS)

But it will be noticed that we get this important result only if (c) and (d) *do* encode different propositions. If they bear the *same* proposition, we have no reason to believe that the referent of a sentence is anything other than the proposition it encodes.

In response to this last point, it is tempting to say:

Of *course* (c) and (d) encode different propositions. You yourself said why. One does, while the other does not, encodes a proposition about hemlock-drinking. Replacing referring terms with co-referring terms often induces dramatic changes in sentence-meaning. "John' s favorite number is smaller than Frank's favorite number" has a very different meaning from "10 is smaller than 11". But "John's favorite number" refers to the same thing as "10", and "Frank's favorite number" refers to the same thing as "11". It would be patent nonsense to insist hat sentence-meaning was always preserved when referring terms are replaced with co-referring terms. In particular, it would be absurd to insist that (c) and (d) have the same meaning.

That is certainly one reasonable response to have.

Here is another very different response.

(c) and (d) both contain definite descriptions. Trivially, definite descriptions either are, or are not, referring terms. Let us consider each case.

If they are referring terms, then it is pretty reasonable to assume that the semantic contribution of "the phi", to the meanings of sentences in which that expression occurs, is the thing that has phi. So if Socrates is that thing, then "the phi smokes" means *Socrates smokes.* If Plato is that thing, then it means *Plato smokes.*

If we grant that "the greatest hemlock-drinking philosopher of antiquity" really *refers* to Socrates, then it is a short-step to saying that the entity semantically contributed by occurrences of that expression is Socrates and Socrates alone. So "the greatest hemlock-drinking philosopher of antiquity smoked" would mean: *Socrates smoked.*

Of course, if this is right, then the expressions

"the wife of the father of the greatest hemlock-drinking philosopher of antiquity"

and

"the wife of the father of the philosopher who figures as the protagonist in most of Plato's dialogues"

both contribute the very same thing; they both semantically contribute Xanthippe (or whoever Socrates' mother was). Consequently (c) and (d) both bear the proposition

(*) *Xanthippe thought her son had bad values.*

Basically, if we assume that definite descriptions are referring terms, then given a very reasonable thesis about reference, it follows that (c) and (d) encode the very *same* proposition. And in that case, we don't yet have any reason to believe that what a sentence refers to is different from the proposition it bears.

It doesn't matter *what* definite descriptions are taken to refer to: *if* "the phi" and "the psi" are co-referential, then "…the phi…" and "…the psi…" mean the exact same thing, at least if we take the plausible view that a referring term semantically contributes its referent. Some have held that definite descriptions are referring terms that refer, not to individuals, but to functions or properties. (Later on we will discuss this position at length.) According to that view,

(*) "the wife of the father of the greatest hemlock-drinking philosopher of antiquity"

and

(**) "the wife of the father of the philosopher who figures as the protagonist in most of Plato's dialogues"

refer to functions or properties of some kind.

For the sake of argument, suppose that is true. In that case, if those terms co-refer, then given a sentence containing the one definite description, replacing it with the other will result in a sentence that means the very same thing as the old one. And if such a replacement results in a sentence that means something *different* from the old one, then the definite descriptions do not co-refer – they refer to different functions or properties.

So our principle seems to be extremely hearty: assuming that the semantic contribution of a referring term is its referent, then “the phi” and “the psi” make the very same semantic contribution if they co-refer. In that case, replacing “the phi” with “the psi” will *not* result in a sentence with a new meaning or sense. Thus, replacing

(*) “the wife of the father of the greatest hemlock-drinking philosopher of antiquity”

with

(**) “the wife of the father of the philosopher who figures as the protagonist in most of Plato’s dialogues”,

will not induce a change in sentence-meaning. In any case, it will not induce such a change *if*, as Frege requires, those terms co-refer.

This would appear to destroy Frege’s argument.

To sum up, if definite descriptions are referring terms, then Frege’s argument does not go through, at least if we hold the plausible view that a referring term semantically contributes its referent.

What if we assume that definite descriptions are *not* referring terms? Might that help Frege’s argument?

Permit me to clarify my meaning. Russell[160] thought of definite descriptions as quantifiers that have no isolable meanings and do not refer to anything. Suppose that Russell is right. Would that help Frege’s argument?

It most certainly would not.

If definite descriptions are *not* referring terms, then (CR) is no longer able to guarantee that (c) and (d) co-refer. If they are *not* referring terms, then (d)

is not what results when a referring term in (c) is replied with co-referring term. And (CR) doesn't say anything about what happens when it is *not* the case that referring terms are replaced with co-referring terms.

Given all of this, it is hard to see how consideration of (c) and (d) could warrant the conclusion that the referent of a sentence is anything different from the proposition it bears. For if (CR) is to guarantee that they have the same referent, then definite descriptions must be treated as referring terms. And *if* they are referring terms, then – being co-referential – the definite descriptions involved semantically contribute the same entity (Xanthippe). In that case, (c) and (d) bear the *same* proposition: which blocks the conclusion that sentence-meaning and sentence-reference come apart.

A Fregean come-back: the semantic contribution of a referring term is its sense, not its referent

It is *not* open to Frege to deny that definite descriptions are referring terms. For if he *were* to deny that, he could no longer use (CR) to get the result he wants.

So he needs definite descriptions to be referring terms. We've seen how this causes problems for his argument.

But our argument makes a big assumption that Frege denies. Our argument assumes that a referring term semantically contributes its referent. And we've seen that *if* a referring term semantically contributes its referent, *then* Frege's argument fails.

But Frege maintained that a referring term semantically contributes its *sense*, not its referent.[161] Permit to explain what this amounts to. Frege held that definite descriptions are terms that refer to individuals. So he thinks that "the inventor of bifocals" refers to Benjamin Franklin. Frege also says

that, when “the inventor of bifocals occurs in a sentence”, it does *not* contribute Benjamin Franklin. Rather, it contributes its *sense*; it contributes descriptive information.

Obviously Frege’s position is not unreasonable, at least not on the face of it.

Frege also held that “Socrates” contributes some kind of descriptive information. Given what Kripke (1972) said, this is not so plausible.

But his position on definite descriptions is plausible (but false, as I’ll argue); it *is*, at least on the face of it, reasonable to say that “the inventor of bifocals” is a referring term whose semantic contribution is a piece of descriptive information.

Of course, he held the same about “the first post-master general”.

If Frege is right about this, then it *is* possible to change the meaning of a sentence by replacing a referring term with a co-referring term. Consider the sentence “The inventor of bifocals snored”. Replace “the inventor of bifocals” with a co-referring description, namely “the first post-master general”. If Frege is right, these two descriptions make different semantic contributions, even though they co-refer. So if Frege is right, then the proposition meant by

(*) The inventor of bifocals snored

is different from that meant by

(**) the first post-master general snored,

even though the one is what results when a referring term in it is replaced with a co-referring term.

So Frege denied the very assumption that, as we just saw, leads to the destruction of his argument. For Frege says: replacing a referring term with a co-referring term *can* change sentence-meaning. In other words, he advocates (SS).

If (SS) is correct, then Frege's gets exactly what he wants, namely: (c) and (d) bear encode different propositions. At the same time, by (CR), they have the same referent.

If we grant Frege (SS), then his argument very much seems to go through. (LL) and (CR) are very hard to deny; presumably, they must be granted. If we *also* grant (SS), then Frege has proven that sentence-reference may come apart from sentence-meaning and, consequently, that sentences do not refer to the propositions that they encode.

For the sake of argument, let us suppose that (SS) holds and, therefore, that sentences do not refer to the propositions that they encode.

The 2nd step in Frege's argument: showing that sentences refer to truth-values

On this basis, Frege is able to show that sentences refer to *truth*-values.

Let S be any true sentence, and let P be any true proposition. *Merely* by replacing referring terms with co-referring terms, we can produce a sentence S* that has P for at least part of its meaning.

Pick any true proposition you wish – say, *arithmetic is incomplete.* And pick any sentence you wish – say (d). Merely by replacing a referring term in (d) with a co-referring term, we can produce a sentence that has the proposition *arithmetic is incomplete* for at least part of its meaning:

(e) "the unique x such that arithmetic is incomplete and x was the father of

the great Greek philosopher who was the protagonist in most of Plato's dialogues had bad value."

This point generalizes without limit. Take any true sentence you wish. Call it S. Take any true proposition you wish. Call it P. If we allow ourselves *only* to replace the referring terms in S with co-referring terms, we can produce a true sentence S* at least part of whose meaning is P.

By CR, S and S* must have the same *referent.*

S and S* needn't bear the same proposition, i.e. they needn't have the same proposition for their meanings. In fact, there is no limit to how much they may differ in respect of what their meanings are.

Indeed, if we allow ourselves only to replace referring terms with co-referring terms, we can destroy *almost* everything that is semantically significant about a sentence.

But the *truth-value* will never changed. That is the one thing that is not destroyed in the process just described.

CR does *not* preserve sense.

CR *does* preserve reference.

Reference is the *only* thing CR preserves.

CR preserves truth-value.

Therefore, sentence-reference must be sentence truth-value. If sentences refer to anything, all true sentences must refer to the truth-value *true.*

An exactly analogous argument shows that all false sentences must refer to the truth-value *false.*

Let us sum up. Given SS and CR, it follows (nearly enough) that all true sentences refer to the same thing – the truth-value *true* – and all true sentences refer to the same thing – the truth-value *false.*

Chapter 3. Our options

There are really only four things we can do. (i) Accept the idea that sentences refer to truth-values. (ii) Reject CR. (iii) Reject LL. (iv) Reject SS.

CR and LL are probably correct. So really it is a choice between (i) and (iv).

What I will do is reject SS.

Why do this? Why not simply *accept* the idea that sentences refer to truth-values?

There are several reasons. Many of these will emerge in due course. Right now I'd like to state just two reasons. The first is: the thesis that sentences *refer* to truth-values runs afoul of some rather reasonable intuitions about reference and semantic content. The second is: if we grant SS, then are forced to grant things which we know aren't true; for example, that there is only one fact and – what is even more counter-intuitive – that there is only true proposition.

I will present the first reason in this chapter, and the second reason in chapter 4.

an anticipatory note

An anticipatory note is appropriate. I will argue (chapter 16) that (LL) is quite ambiguous. We say that the following expressions are "logically equivalent":

"Plato"

and

“the unique thing x such that 1+1+2 and x is Plato”

We also say that the *following* two expressions are “logically equivalent”:

“triangles have three sides”

and

“1+1=2”.

We will find that the term “logical equivalence” denotes completely different relations in the two cases. The “logical equivalence” that holds between the first two expressions is as distinct as can be from the “logical equivalence” that holds between the second two.

We will also find, I believe, that because the term “logically equivalent” is thus ambiguous, at least *some* slingshot-style arguments are invalidated (though not, I believe, all of them).

But I want my main case against the slingshot *not* to involve any problems supposedly attaching to (LL). I want to *grant* (LL), in all its glory, to advocates of the Slingshot, and show that *even if* we make that concession, the Slingshot still wobbles.

So, in what follows (until the end of this book), it will be assumed that (LL) is correct, and that the Slingshot’s reliance on it does *not* vitiate it.

Now I will present my two main reasons for thinking that something must be wrong with the Slingshot.

The first reason to question the Slingshot

Consider the sentence "George W. Bush smoked a cigarette on April 2, 2005". You and I both understand that sentence about as well as it can be understood. We completely and thoroughly grasp its semantics. But we don't have the slightest idea whether it is true or not.

Except perhaps in very special cases, knowledge of the semantics of a sentence absolutely does not require knowledge of its truth-value. (The special cases I have in mind are analytic propositions like "Socrates is identical with Socrates". I'm not even convinced that a grasp of the semantic content of *these* sentences necessarily involves a grasp of their truth. And even if it does, that has nothing to do with semantics; it has to do with epistemology.)

Given this, it is hard to see how the semantic value of a sentence could be a *truth*-value. One can know everything there is to know about the semantics of a sentence – about its meaning -- without knowing anything about its truth-value. So it is hard to see how the truth-value of a sentence could be any part of its *semantics.* Therefore it is hard to see how the truth-value of a sentence could be its referent. For if a term refers, its referent is an integral part of its semantic content.

Suppose that "Blark" is somebody's name, but you have no idea who that person is. So "Blark" refers to someone; you just don't know who. Because you don't know who "Blark" refers to, you really don't know its semantics. You don't know what proposition is semantically encoded in sentences like "Blark is a nice guy". Of course, you know that that sentence will be true if *somebody* named "Blark" is a nice guy. But the proposition *somebody named 'Blark' is a nice guy* is completely different from – doesn't entail, and isn't entailed by – the proposition *Blark is a nice guy.*

If a term refers, then you simply don't know its semantics until you know who or what it refers to. Granted, you can have a *partial* knowledge of the semantics of "Blark" without knowing who or what its referent is. If you know that it is a name, then perhaps you have *some* knowledge of its semantics. But you are still quite ignorant of an important part of its semantics – of what is, really, the essential part of its semantics.

If sentences are terms that refer to truth-values, then you certainly couldn't have a complete grasp of their semantics unless you knew what their values were. But typically one *can* have a complete grasp of the semantics of a sentence without knowing its truth-value. (By contrast, one *cannot* have a complete knowledge of the semantics of a sentence without knowing what proposition it encodes. This fact will constitute a part – a small part – in my argument that sentences refer to propositions.) It seems to be of the essence of language that one should be able to grasp the *meanings* of sentences, and sentence-tokens, without knowing their truth-values. The idea that sentences semantically contribute truth-values puts a serious strain on this idea.

There is an obvious response to all of this:

Consider the definite description "the richest man in Holland". You can understand it perfectly without having the foggiest idea to whom it refers. Consider the expressions: "the combination of chemicals that cure AIDS", "the sequence of sentences that proves the most important mathematical result of the 23rd century", "the exact number of people in Nepal". You have no idea what these expressions refer to. But you understand them perfectly. So one *can* grasp the semantics of an expression without knowing its referent. So given only that one can grasp the semantics of a sentence without knowing its truth-value, it does not follow that a sentence's referent is anything other than its truth-value.

This is an important objection. But it is based on a major fallacy. My response to that objection is complicated. So let me outline before giving it in full.

My response: an outline

"The richest man in Holland" either does, or does not, refer to a certain individual. If doesn't, then it is a quantifier. In that case, it doesn't refer to anybody. And in that case, to grasp the semantics of that expression is to know what its quantificational structure is. You'd have to know that "… the richest man in Holland…" means: *someone x is uniquely a richest man in Holland and…x…*

So if definite descriptions do not refer to individuals, then you absolutely *do not* know the meaning, the semantics, of "the richest man in Holland" unless you know what quantifier it is. So if the objector's point is to go through, then definite description must refer to individuals.

Very well. Let's suppose that Russell is wrong. In that case, "the richest man in Holland" has a two-dimensional semantic structure. There is what particular *tokens* of it refer to. And there is also a general rule, a function, pairing off tokens of it with a certain individual (the individual who satisfies the open-proposition: *x is a unique richest man in Holland*). That general rule is the semantic content of the expression-*type* "the richest man in Holland". The semantic content of that *type* is absolutely not some individual; it is, to repeat, a certain function.

Now you absolutely do *not* know the semantics of that expression-*type* unless you know that it is associated with that particular function. But that

expression *type* doesn't refer to anything; in any case, it doesn't refer to any individual; it refers to a function, if it refers to anything. And you couldn't possibly grasp that expression-*type's* semantics without knowing that it expressed, or denoted, that function. So with regard to our principle that one must the know the referent of a referring term to grasp its semantics, that I principle is thus far upheld.

What about *tokens* of "the richest man in Holland"? Supposing that "the richest man in Holland" is not to be given a Russellian analysis, then the semantic content of *tokens* of it (right now) is (let us say) Jan van Eykel. But one surely couldn't know the semantic content of a *token* of "the richest man in Holland" unless one knew that it referred to Jan van Eykel.

No matter how you slice it, *if* an expression is a referring term, then you really can't know much about its semantics unless you know what it refers to. But, plainly, one can typically know everything about the semantics of a sentence without knowing what it refers to. In fact, this seems to be essential to the very concept of a sentence: a sentence must be something that we understand without knowing its truth or falsity. If we knew the truth or falsity of every sentence, then language wouldn't be capable of transmitting *new* information. So the truth-value of a sentence cannot part of its semantics, and thus cannot be its referent.

a more complete version of this argument

Let me give a less summary exposition of the view just stated.

People disagree about the semantics of definite descriptions. Some people say they are terms that refer to individuals ("the inventor of bifocals" refers to Benjamin Franklin). Others say they are terms that refer to functions or

properties. Yet others say that they don't refer to anything at all (Russell).

No matter which of these analyses is the correct one, it turns out that one *does* have to know what a referring term refers to, if one is to know its semantics.

Suppose that definite descriptions are expressions that refer functions. So "the inventor of bifocals" refers to a function. First of all, which function would it be? Supposing that "the inventor if bifocals" does not refer to individuals, then "…the inventor of bifocals…" means: exactly one thing invented bifocals and…that thing…

So "the inventor of bifocals snored" is true just in case the class of snorers contains exactly one member that is a bifocal inventor and nothing falling outside of that class is a bifocal inventor. Thus "the inventor of bifocals" can be thought of as a function that assigns *truth* the class of snorers iff the class of snorers contains exactly one member that is a bifocal inventor and nothing falling outside the class of snorers is a bifocal inventor. In general, "the inventor of bifocals" is an expression that refers to a function that assigns truth to a class C iff C contains exactly one bifocal inventor and nothing falling outside of C is a bifocal inventor.

If that is right, then one hasn't the faintest knowledge of the semantics of "the inventor of bifocals" unless one knows that it refers to that function.

We get the same result if we suppose that it refers to some second-level property.

Once again, if we suppose that "the inventor if bifocals" does not refer to individuals, then "…the inventor of bifocals…" means: *exactly one thing invented bifocals and…that thing…*

Thus "the inventor of bifocals snored" is true iff the class of snorers has the following property: *anything falling in that class is a bifocal inventor and nothing falling outside it is a bifocal inventor*. In general, "the inventor of

bifocals" can be seen as denoting a property of classes: the property of being such that something falling in it is a unique bifocal inventor.

If "the inventor of bifocals" does indeed denote that property, then one hasn't the faintest knowledge of the semantics of "the inventor of bifocals" unless one knows that it denotes that property. Consequently there is no knowledge of the semantics of a definite description unless there is knowledge of its referent. And our principle – no knowledge of the semantics of a referring term without knowledge of its referent – is upheld.

Now suppose that definite descriptions don't refer to anything. In that case, vacuously, it is impossible to know to know the semantics of a definite description without knowing its referent. So our principle continues to be upheld.

Finally – this is the really interesting case - suppose that definite descriptions are what they appear to be: expressions that refer to individuals. So, in this world, "the inventor of bifocals" refers to Benjamin Franklin. In other worlds, it refers to Jefferson.

Given this, let w be our world and let w* be some other world where English is spoken. Suppose that, in w*, Jefferson invented bifocals.

By supposition, "the inventor of bifocals" is a term that refers to individuals. Also by supposition, the expression-type "the inventor of bifocals" has the exact same semantics in w* that it has here. Finally, by supposition, in w* "the inventor of bifocals" refers to Jefferson.

These suppositions are all coherent; and they are not tendentious.

But we have a problem. By supposition, "the inventor of bifocals" has the same *semantics* in both w and w*. Presumably, the referent of a referring term is at least part of its semantic content. "The inventor of bifocals" refers to one thing in w and to a different thing in w*. Thus, it would appear, "the inventor of bifocals" doesn't have quite the same semantic content in w* that

it has in w.

But by assumption it *does* have the same semantics in both worlds.

We don't want to say that the referent of an expression is *not* a part of its semantic content. That would be absurd. If the referent of a referring expression isn't a part of its semantics, what is?

But we *do* want to say that "the inventor of bifocals" refers to different things in different worlds, even though its semantics stays the same.

There is an easy way to make sense of all this. Strawson[162] showed us the way.

We begin by distinguishing the expression-*type* "the inventor of bifocals" from *tokens* of that expression. Further, we say that the expression-*type* doesn't refer to anything. But the expression-type does have meaning. The meaning of that expression-type is a rule that assigns a referent to *tokens* of it, depending on the context of tokening. The world in which a tokening occurs is, at least in this case, part of the context.

So the meaning of the expression-*type* "the inventor of bifocals" is some rule that is at least vaguely like the following. Given any token of "the inventor of bifocals", if that token occurs in a context where there is exactly one unique bifocal inventor x, then that token refers to x.[163]

Thus, in w, tokens of "the inventor of bifocals" refer to one thing (Benjamin), while in w*, they refer to a different thing (Jefferson). For a difference in worlds means a difference in contexts of utterance (where this class of expressions is concerned).

So the semantic content of a *token* of "the inventor of bifocals" in w is different from the semantic content of a *token* of that expression in w*.

At the same time, the meaning of the expression *type* "the inventor of bifocals" is the same in both w and w*. But the expression *type* doesn't refer

to anything; the meaning of that type is some rule that assigns individuals to tokens of that type.

Now we can close the argument. One cannot know the semantic content of a *token* of "the inventor of bifocals" unless one know who the referent of that token is.

Now one *can* know the semantic content of the expression-*type* "the inventor of bifocals" without knowing what it refers to. But that is only because it doesn't refer to anything. *Tokens* of it refer; the expression-type does not.

Thus, whenever we have something that refers, it turns out that one *does* have to know its referent to knows it semantics.

An analogy may help clarify what I am trying to say.

The word "you" can be used to different things. But its meaning stays the same. When I say "you are tired" to Jerry, the word "you" has the same meaning it has when I say "you are drunk" to Fred. It isn't as though it means one thing when I address Jerry, another thing when I address Mary, a third thing when I address Ethel.

So we seem to have a word with a variable reference and a constant meaning.

One way to deal with this is to say that the referent of a word is not a part of its meaning. But that would be absurd or at least very arbitrary. Presumably, the meaning of "Plato" is its referent. (Of course, this presumption has been challenged. But it is has become increasingly hard to deny. And we ourselves will substantiate it). If we say that the referent of "you" is *not* even a part of its meaning, then we run the risk of rendering "meaning" quite ambiguous.

In any case, we don't have to say that "you" has a constant meaning and variable reference. There is a better way to approach.

We say that *tokens* of "you" refer to different things, and that the *type* doesn't refer to anything. The semantic content of the type is a rule ("character") which assigns referents to the tokens.

You obviously can*not* know the semantic content of a token of "you" (or any other indexical) unless you *do* know what it refers to. If the *type* "you" refers to anything, it isn't to individuals; it is some to rule, some function, that pairs off tokens of that expression with individual. And, obviously, if you don't know that *rule*, then once again you have not the foggiest idea of what the expression-type "you" refers to. So, where indexicals are concerned, there is plainly no semantic knowledge without knowledge of expression-reference.

If definite descriptions are expressions that refer to individuals, then they have an analogous, two-dimensional structure. There is the semantic content of *tokens* of "the inventor of bifocals"; this is the thing they refer to. And there is also the semantic content of the *type* "the inventor of bifocals", that being some rule that pairs off the just mentioned tokens with some individual. So, for reasons strictly analogous to those just given in connection with indexicals, there is plainly no knowledge of the semantics of definite description tokens or definite description-types without knowledge of what, if anything, they refer to.

I should close this discussion by dealing with an objection:

Consider the definite descriptions "the first whole number that comes after four". This definite description refers to the same thing in all worlds. It seems strange to say that the expression-*type* doesn't refer to anything. For plainly it does – the number five. The distinction between expression-type and expression-token seems to make sense only where there is the possibility of variable reference. Obviously "you" has a variable reference, but not a

variable meaning. So we distinguish the expression itself from tokens of it; and we use two-dimensional semantics to make sense of it all. But "the number that comes after four" does *not* have a variable reference; its referent is *not* dependent on the context of utterance. A two-dimensionalist approach is warranted only where reference is a function of context of utterance. So a two-dimensionalist approach simply isn't appropriate when it comes to "the number that comes after four". The same is true of any definite description that rigidly designates. Since "the whole number that comes right after four" doesn't have a two-dimensional structure, it seems to follow that "the inventor of bifocals" doesn't either. So your analysis is wrong.

It may happen – for reasons of metaphysics, logic, or theology – that a certain description denotes the same thing in all worlds. But that is not the concern of semantics. It is patently obvious that *semantically* "the inventor of bifocals" is in the same class as "the whole number that comes right after four". It is also clear, given what we just said, that "the inventor of bifocals" must be taken to have a two-dimensional structure (unless it is treated as a quantifier). So the natural course to take is to say that "the whole number that comes right after four" also has a two-dimensional structure but that, for reasons of mathematics that are totally irrelevant to semantics, tokens of that expression always refer to the same thing.

The *semantics* of "the whole number that comes right after four" leave it open whether it refers to the same thing all worlds. The *mathematics* involved don't leave it open. But when we say that the referent of an expression is "dependent" on the context of utterance, we are talking about *semantic* dependence; we are saying: given only the semantics of that expression, it is open what it refers to in different context.[164] It might be that for logical or

metaphysical or theological reasons, it always refers to the same thing. But those reasons lie outside the scope of semantics.

We don't want our semantic analyses to prejudge logical or metaphysical issues. I myself believe that (tokens of) "the whole number that comes right after four" refer(s) to the same thing in all worlds (I mean: in all worlds where it has the semantics that it actually has, it refers to the same thing.) But Quine and Putnam have argued that all truths are empirical – that *five is the number that comes after four* is empirical[165] - and, consequently, that there are epistemically possible worlds where six is the thing denoted by "the whole number that comes right after four". I myself think this is ridiculous. But surely we don't want our *semantic* analyses to prejudge this issue. We don't want to say: "given that 'the whole number that comes right after four' rigidly designates, we must reject a two-dimensionalist approach to definite descriptions". For, conceivably[166], a Quine or a Putnam might prove that, in some worlds, the number six is the thing referred to by "the whole number that comes right after four".

summary

If an expression-type or expression-token refers, one must know what it refers to know its semantics – to know what it means. Obviously one can know what sentences, and sentence-tokens, mean without having any idea whether they are true or false. So whatever sentence-reference is, it isn't truth-value. And it must be something that a person knows in virtue of knowing the semantics of sentences (or sentence-tokens, as the case may be): which makes propositions be the obvious candidate.

an objection

You say that one cannot know the meaning of referring term without knowing its referent. This seems to be contradicted by certain recent developments in semantics and epistemology. One can understand the word "water" without knowing that it refers to H_2O. One can understand the word "Hesperus" without knowing that it refers to Phosphorous.

One cannot understand the word "water" without knowing what it refers to. It refers to water; and if you don't know that, you don't know what it means. Of course, you can know that "water" refers to water without knowing everything there is to know about water – without, for example, knowing that it has a certain chemical composition. But that is irrelevant. To know *what* a term refers to is not to know everything fact about that term. To be acquainted with a thing is not to know everything about it.

To know what "Hesperus" means, one must know that it refers to Hesperus. One doesn't always *recognize* objects with which one is acquainted. One may not *recognize* Hesperus in a certain celestial body that one is seeing. A consequence is that one may use different names to refer to the same thing. (If we always *recognized* objects with which we were acquainted, we would no longer unknowingly refer to the same object with different names.)

But none of this undermines the point that one does have to know what "Hesperus" refers to, if one is to understand it. The fact that we do not recognize Hesperus every time we encounter it reduces to the banal fact that we do not typically[167] know everything there is to know about the objects with which we are acquainted. Plainly, if "Hesperus" is a referring term, then

I must be able to associate it with the right object if I am to understand it.

The objection just made is an important one, and shouldn't be glibly dismissed. But surely the thing to say is not: *One can know what "water" or "Hesperus" mean without knowing what they refer to.* That is practically a self-contradiction. What we must say is this. We are acquainted with various things. Given such a thing, we don't always recognize it. So I can know what a certain term refers to, without always recognizing a thing *as* that referent; given only that I am acquainted with Hesperus, it doesn't follow that I always recognize Hesperus *in* the thing I see as I gaze upon the heavens.

an application

An old saw in the philosophy of language is: there is a difference between meaning and reference, and between intension and extension.

The distinction between meaning and reference is a confusion. So far as it is legitimate, it is really a distinction between the meaning of expression-*types* and the meanings of expression-*tokens*.

First of all, the meaning-reference distinction obviously doesn't arise with words that are mere labels, e.g. "Socrates". It arises only with referring terms that have a descriptive component, e.g. indexicals and (perhaps) definite descriptions. "The inventor of bifocals" has a descriptive component; because, as matter of semantics, if a person is to be picked out by (tokens of) it, that person must satisfy a certain description. So "the inventor of bifocals" has meaning and reference; and these are different.

But that is spurious. There is never any *one* expression that has both meaning and reference. The *type* "the inventor of bifocals" has no referent; only tokens of it do. The *type* has a rule for its meaning, and is descriptive. *Tokens* of that expression have no descriptive content; they have only

individuals (or nothing) for their meanings.

Exactly similar remarks apply to indexicals and, indeed, to all other expressions that seem to have both meaning and reference. [168]

Summary

You can know the semantics of a sentence-types, and of sentence-tokens, without knowing whether they are true or false. I can understand a token of "that guy is an imbecile" without knowing whether it is true or false. And, of course, I cannot understand the meaning of that sentence-type without knowing what *its* truth-value is; for that sentence-type does not have truth-value. If sentence-types or –tokens referred to truth-values, then we couldn't grasp their semantics without knowing their truth-values. So they don't refer to truth-values.

another argument that sentences do not refer to truth-values

Here is one way of arguing the point just made, or one similar to it. Consider the sentence "George W. Bush smoked a cigarette on April 2, 2005". Let w be our world, and suppose that in w that sentence is true. Let w* be a world just like ours – so people speak English, and English is characterized by the same semantics as in w – but George W. Bush did not smoke a cigarette in w* on April 2, 2005.

It is a given that "George W. Bush smoked a cigarette on April 2, 2005" has the exact same semantics in both w and w*. This, it seems, is tantamount to saying that it has the same semantic content, makes the same semantic

contributions, and so on. The problem is that “George W. Bush smoked a cigarette on April 2, 2005” is true here and false in w*. So if that sentence is true in this world and false in that, then it refers to and, therefore, semantically contributes, *different* things. But it is incontrovertible that, in both w and w*, that sentence has the very same semantics. So to keep alive the idea that truth-value is reference, we must somehow sever the notion of reference from the notion of semantic-contribution. Such a move, it seems to me, borders on incoherence.

Notice that, in both w and w*, “George W. Bush smoked a cigarette on April 2, 2005” expresses the same *proposition*. So it has precisely the same semantics – makes the same semantic contribution – and expresses the same proposition. The idea that sentences refer to propositions is more consistent with this datum than the idea that they refer to truth-values.

Here is an objection:

Some connectives are truth-functional. “S and S*” is true iff both S is true and S* is true. The truth of the conjunction is a function of the truth-values of its components. So surely the truth-values of S and S* are at least among the things which they semantically contribute in that context. In general, it seems that in some contexts, what a sentence semantically contributes is a truth-value. So you’re wrong to say that sentences don’t semantically contribute their truth-values.

Here is my reservation about this. Let S be the sentence “George W. Bush smoked a cigarette on April 2, 2005” When it occurs on its own – if I just blurt out that sentence -- it is occurring in a truth-functional context, no less than when it is occurring in the scope of a truth-functional operator like

"and". But, for reasons we've seen, one can grasp the semantics of that sentence without knowing its truth-value.

Obviously "and" is truth-functional. But this doesn't mean that sentences falling in its scope semantically contribute their truth-values. Under no circumstances are the truth-values of S and S* part of what they *semantically* contribute to "S and S*". If they *did* semantically contribute their truth-values, one couldn't even *understand* (grasp the semantics of) the conjunction without knowing those truth-values.

"S and S*" is equivalent to: the sequence <S, S*> is such that both its members are true. So "and" can be thought of as a property of sequences of sentences or propositions. Let CN be that property. CN is truly attributed to some sequences and falsely attributed to others, just as the property of being green is truly attributed to some individuals and falsely attributed to others. The truth of "Kermit is green" depends, of course, on whether Kermit has the property of being green. But the semantic contribution of "Kermit" is not the property of being green. The truth of "<S, S*>" depends on whether S and S* have certain truth-values. But they no more semantically contribute those truth-values than "Kermit" semantically contributes the property of being green in "Kermit is green".

To sum up: one can fully grasp sentence-semantics without knowing sentence-truth-value. So truth-value cannot be a part of sentence-semantics. The referent of a referring expression is the essence of its semantics. So a sentence's truth-value cannot be its referent. Its referent must be something that one grasps in virtue of understanding the sentence.

Chapter 4. *the second reason to question the slingshot: it proves too much*

The second reason *not* to accept that sentences refer to truth-values is that

Frege's argument for this conclusion proves too much; it proves things that aren't true. Therefore that argument must involve a fallacy.

The slingshot, I will now show, proves there is only one fact and also (what is more alarming) that there is only one truth; it also proves *either* that "necessarily" is truth-functional *or* that sentences falling in its scope don't have their customary meanings (so "1+1=2" means one thing on its own, and another in "necessarily 1+1=2").

Since the slingshot has these disastrous consequences, we must conclude that it involves a fallacy. (For reasons stated earlier, the only possible mis-step in the fallacy is (SS).)

Let me now say why the slingshot has these consequences.

Davidson's slingshot

Donald Davidson used a version of the slingshot to show that there is only one fact. (He accepted this conclusion.[169] And, on the basis of it, he drew some far reaching conclusions about semantics, the theory of knowledge, and the nature of mind.[170])

Consider the following sequence of expressions::

(i) the fact that snow is white.
(ii) that fact that [the class of all things x such that such that (x=x and snow is white) is identical with the class of all things x such that (x=x)].
(iii) the fact that [the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x)].
(iv) the fact that [Grass is green].

Davidson thinks this shows that all these expressions refer to the same *fact.* Why does he thinks this?

Here is his reasoning.

First of all, (i) and (ii) are logically equivalent. So (LL) guarantees that (i) and (ii) do not refer to *different* things. So they don't refer to different facts.

(CR) guarantees that (ii) and (iii) don't refer to different things or, therefore, to different facts.

Finally, (LL) guarantees that (iii) and (iv) don't refer to different things or, therefore, to different facts.

According to Davidson, this shows us that "the fact that grass is green" and "the fact snow is white" refer to the same fact. Of course, given *any* two expressions of the form "the fact that phi" and "the fact that psi", an analogue of the argument just given shows that they refer to the same fact. So the fact that Mozart was a genius is identical with the fact that the moon is not the shape of a cube.

Davidson's argument certainly *seems* compelling. But he makes a couple of unstated assumptions. First, he assumes that definite descriptions are referring terms. If they are not referring terms, then CR doesn't guarantee the step from (ii) to (iii).

Davidson *also* assumes (SS). Let me explain why. Let Omni be the class of all objects x such that (x=x).[171] If the semantic contribution of a referring term is *just* the thing it refers to, then

"the class of all things x such that such that (x=x and snow is white)"

contributes the very same thing as

"the class of all things x such that such that (x=x)" .

Both expression pick out Omni, and Omni alone.

Consequently, both (ii) and (iii) end up being synonymous with: 'the fact that Omni is Omni". Thus the core of Davidson's argument becomes:

(a) the fact that snow is white.
(b) that fact that Omni is Omni.
(c) the fact that Omni is Omni.
(d) the fact that Grass is green.

In other words, (i) ends up being synonymous with (a), (ii) with (b), and so on. But, in that case, Davidson's argument is a failure. For (LL) no longer guarantees that (i) and (ii) refer to the same thing, or that (iii) and (iv) refer to the same thing.

To block this destruction of his argument, Davidson needs

" the class of all things x such that such that (x=x and snow is white)"

to *not* make the same semantic contribution as

"the class of all things x such that such that (x=x)" .

But he *also* needs both those expressions to be co-referring expressions; otherwise he can't use (CR) to guarantee that (ii) and (iii) co-refer.

So Davidson needs it to be possible to change the sense of a sentence by replacing referring terms in it with co-referring terms. In other words, Davidson needs (SS).

Clearly once (SS) is granted, then Davidson is right: (i) and (iv) *cannot* refer to different things or, therefore, to different facts. It follows that *any* two expressions of the form "the phi" and "the psi" co-refer. From this, in turn, it follows (nearly enough, though probably not quite[172]) that there is only one fact.

The main point is: once we grant (SS), and that definite descriptions are referring terms, the way is clear to proving that there is only one fact.

An analogue of Davidson's slingshot shows that any two true propositions are logically equivalent.

(i_L) that snow is white.
(ii_L) that [the class of all things x such that such that (x=x and snow is white) is identical with the class of all things x such that (x=x)].
(iii_L) that [the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x)].
(iv_L) that [Grass is green].

Once we grant (LL), (CR), and (SS), along with the idea that definite descriptions are referring terms, the argument just given goes through. In other words, if we grant the very assumptions that make Frege's argument work, then we can show that any two propositions are logically equivalent.

Indeed, an analogue of Davidson's original argument shows that there is only *one* true proposition:

(i_P) the proposition that snow is white.
(ii_P) the proposition that [the class of all things x such that such that (x=x

and snow is white) is identical with the class of all things x such that (x=x)].
(iiiP) the proposition that [the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x)].
(ivP) the proposition that Grass is green.

For reasons exactly analogous to those just given, it follows that if (SS) holds, and if definite descriptions are referring terms, then there is only one proposition.

This result is catastrophic.

Once we grant (SS), and once we grant that definite descriptions are referring terms, then the way is clear to proving that there is only one proposition: and thus only one truth and one falsehood. I know every true proposition since I know one true proposition. I know everything (at least everything that can be embodied in a proposition) since I know my own phone number.

Of course, the two assumptions just described are *exactly* what Frege needs to show that sentences refer to truth-values. Remember what we said earlier. If (LL), (CR), and (SS) are true, *then* it follows that sentences refer to truth-values. Recall also that (LL) and (CR) are beyond reproach. The only one that can be questioned is (SS). We've seen that, if (SS) is granted, then it follows that there is only one true proposition. But obviously there is more than one proposition. So (SS) must be false. But if (SS) is false, then Frege's argument is fallacious.

The slingshot proves *too* much. Of all the premises involved, the only questionable one is (SS). So (SS) must be wrong.

a Quinean slingshot

I would like to give another reason to think that the slingshot proves *too* much and that, consequently, (SS) must be wrong.

Some sentence-connectives are “extensional”. The truth of “snow is white and grass is green” is wholly a function of the truth-values of the sentences falling within the scope of the “and”. So “and” is extensional. Some sentences very much appear *not* to be extensional. The truth of “Khrushchev resigned after Kennedy died” is *not* a function of the truth-values of the constituent sentences (“Kennedy died”, “Khrushchev resigned”). Both those sentences are true; but that is compatible with “Khrushchev resigned after Kennedy died” being false.

One point should be made clear. Orthography doesn’t always make it clear what the *scope* of a sentence connective is. In the sentence “Grass is green and snow is white”, the *scope* of the “and” is really the (ordered) pair <grass is green, snow is white>.[173] English-notation doesn’t make this fact clear. (So-called “Polish” logical notation does make it clear.) The “and” can be thought of as a function that takes that pair as an argument and pairs it off with another proposition (a conjunctive proposition). But what is important is that *both* sentences fall within the scope of the “and”.

The same thing is true of “after”. In the sentence “Khrushchev resigned after Kennedy died”, *both* constituent sentences fall within the scope of the “after”, even though only one of those sentences is to the right of it.

Two more pieces of terminology must be defined. Let us say that a connective is PSLE[174] iff there is no change in truth value as a result of replacing a sentence falling in its scope with a logically equivalent sentence. A clear example of such a connective is “necessarily”.

Finally, let us say that a sentence-connective "induces opacity" iff sentences falling with its scope do not have their ordinary meaning. "And" does not induce opacity; for "snow is white" has the same meaning in "snow is white and grass is green" that it has on its own. I cannot give an uncontroversial example of a connective that induces opacity. I myself don't think that *any* sentence-connectives induce opacity. That is actually one of the things I am trying to show. But Frege thought that epistemic connectives – e.g. "John believes that" -- induced opacity.

A slingshot-style argument can easily be constructed that shows that at least *one* of the following two propositions is true:

(TF) All PSLE sentence-connectives are truth-functional.
(OP) All PSLE sentence-connectives induce opacity.

Let "Phi" be an arbitrary PSLE connective. And let S and S* be any two sentences that have the same truth-value.

TF

(1) Phi (S)".
(2) Phi ([x:x=x and S]=[x:x=x])
(3) Phi ([x:x=x and S*]=[x:x=x])
(4) Phi (S*).

Given that "Phi" is PSLE, we clearly have safe passage from (i)-(iv). It seems to follow that *any* PSLE connective is truth-functional.

But this is demonstrably not the case. Let "Phi" be "is metaphysically necessary". Let S be *1+1=2* and let S* be *Smith is bald*. TF shows *either* that *Smith is bald* is metaphysically necessary *or* that "necessarily" induces

opacity. Obviously *Smith is bald* is not metaphysically necessary. So we must say that "Necessarily" induces opacity. Thus, "1+1=2" means one thing on its own, and a completely different thing in "necessarily 1+1=2", and yet a different thing in "it is necessarily the case that necessarily 1+1=2".

Some hearty spirits have maintained that "necessarily" induces opacity. But it is not a plausible position. And, I think, it is demonstrably false. Let "necessarily*" have the same meaning as "necessarily", except that, by our stipulation, any sentence S falling within the scope of "necessarily*" bears the proposition that it customarily (in truth-functional contexts) bears. There doesn't seem anything wrong with this connective. If I say "necessarily* 1+1=2" I am saying, of the proposition customarily borne by "1+1=2", that it is necessary. Surely there is some proposition to just that effect.

An analogue of TF shows that either *Bob is bald* is necessary* *or* that "necessarily*" induces opacity. We know the latter is false. And we know, on independent grounds, that the former is false. So TF must be fallacious.

Here is another example of a connective that shows TF to be fallacious. The fact that grass is green is causally explained, at least in part, by the fact that grass contains chlorophyll. Let EXP be a two-place sentence-connective that we define as follows. "EXP (S, S*)" is true just in the following holds. Let P and P* be the propositions *customarily* borne by S and S*. Suppose that P and P* are both true. Either P*, or one of the propositions composing it, provides at least a partial explanation of either P or one of the propositions composing it So

(e1) EXP (grass is green, grass contains chlorophyll)

is true. Plainly "EXP" is PSLE. For if we replace either constituent sentences with a logically equivalent sentence, what results is plainly true, given how

we defined “EXP”:

(e2) EXP (grass is green, 1+1=2 and grass contains chlorophyll).

Is true. And so is

(e3) EXP (grass is green and 1+1=2, grass contains chlorophyll).

Given how we’ve defined, it is pretty clear that “EXP” does not induce opacity. After all, it is the essence of “EXP” that sentences falling in its scope bear their *customary* meanings.

At the same time, PSLE is plainly *not* truth-functional.

(e4) EXP (grass is green, Mozart was a great composer)

is obviously false. The fact that Mozart was a great composer does not, by any stretch, have any explanatory relevance to the fact that grass is green. And I’m pretty sure that everyone – even Quine, who puts forth TF (or, at least, a very similar argument) – would agree.

An advocate of TF might say that EXP *does* induce opacity, despite what we’ve said.

But this is not a promising measure, given how “EXP” is defined. Also, even if “EXP” did induce opacity, that wouldn’t automatically mean that (e4) was true. We’d need some guarantee that the propositions borne by the two sentences in question *after* opacity was induced were such that one of them was explanatorily relevant to the other. But no such guarantee appears to be forthcoming.

Another tack an advocate of TF might take is to say that nothing is

explanatorily relevant to anything. But this would obviously be desperate. And, even if true, it wouldn't really speak to the larger issue. For we could obviously construct temporal or moral analogues of "EXP"; and the explanatory nihilism just stated wouldn't have any relevance to these.

The natural, and virtually inescapable, conclusion is to say that TF is fallacious. It is *not* true that PSLE connectives are always truth-functional.

The question is: what is the problem with TF? The problem, I submit, is that it (covertly) relies on SS.

Let us start with some uncontroversial points. (2) contains the definite descriptions: "the class of all things x such that x=x and S". (3) contains the definite description: "the class of all things x such that x=x and S*". Both contain the expressions "the class of all things x such that x=x".

TF assumes that "S" is logically equivalent with "([x:x=x and S]=[x:x=x])". It also assumes that "S*" is logically equivalent with "([x:x=x and S*]=[x:x=x])".

Trivially, either definite descriptions are referring terms or they aren't. Let's consider each case. First, let's suppose that they are not referring terms. In that case, they are Russellian quantifiers. But then TF fails miserably. The transition from (2) to (3) no longer goes through. Let S be the sentence "Bach wrote fugues". Let S* be the sentence "Stalin was evil". In that case, (2) and (3) become:

(2_R) There is some class C such that, for any x, x falls into C iff (x=x and Bach wrote fugues), and there is some class C* such that, for any y, y falls into C* iff (y=y), and C=C*.

(3_R) There is some class C such that, for any x, x falls into C iff (x=x and Stalin was evil), and there is some class C* such that, for any y, y falls into

C* iff (y=y), and C=C*.

(2_R) and (3_R) are definitely *not* logically equivalent. So the fact that "Phi" is PSLE does *not* warrant the transition from (2) to (3), at least not if definite descriptions are read as Russellian quantifiers.

One might say:

Obviously (2_R) and (3_R) are not logically equivalent. But this doesn't mean that the transition from (2) to (3) fails. It means that it fails *only* if "Phi" is given wide-scope, i.e. only if (2_R) and (3_R) in their entirety fall within the scope of "Phi". But if "Phi" is given narrow-scope, then the transition from (2_R) to (3_R) works. There are different narrow-scope readings. Perhaps TF goes through on one of them.

Pick any narrow-scope reading you wish. The result will *in*validate one of the steps in TF. Here is the most obvious narrow scope reading of (2) and (3).

(2_R) There is some class C such that, for any x, x falls into C iff (x=x and Bach wrote fugues), and there is some class C* such that, for any y, y falls into C* iff (y=y), and Phi (C=C*).

(3_R) There is some class C such that, for any x, x falls into C iff (x=x and Stalin was evil), and there is some class C* such that, for any y, y falls into C* iff (y=y), and Phi (C=C*).

Let us put this reading into the context of the argument as a whole:

(1) Phi (Bach wrote fugues)
(2_R) There is some class C such that, for any x, x falls into C iff (x=x and Bach wrote fugues), and there is some class C* such that, for any y, y falls into C* iff (y=y), and Phi (C=C*).
(3_R) There is some class C such that, for any x, x falls into C iff (x=x and Stalin was evil), and there is some class C* such that, for any y, y falls into C* iff (y=y), and Phi (C=C*).
(4) Phi (Stalin was evil).

In that case, the steps from (1) to (2), and from (3) to (4), become completely invalid. For any objects C and C*, the proposition *Bach wrote fugues* is not equivalent to the proposition *C=C**. The same is true of *Stalin was evil.* So if we give "Phi" that particular narrow scope reading, then two much needed transitions simply fail; they are no longer guaranteed by the fact that "Phi" is PSLE.

Granted, there are other narrow-scope readings. But it is easy enough to verify that, on any of them, one of the steps in question becomes invalid. (It would be a tedious and unnecessary exercise to go through each one of them.)

It is clear what we must say: we must treat the definite descriptions involved – "the class of things x such that x=x", and so on – as *referring* expressions. If we *don't* do this – if we treat them as quantifiers – then TF becomes a tissue of non-sequiturs.

Here is the problem. For the sake of argument, let the definite descriptions in TF be referring expressions, And, once again, let Omni be the class of all things x such that x=x. In that case, all of the definite descriptions in TF co-refer: they *all* refer to the very same class --- Omni. Granted, they refer to it in different ways. But there is some *one* class that they all pick out.

If, as we are supposing, definite descriptions are referring terms, then it is natural to suppose that the sole semantic contribution of such an expression is the thing it picks out. If that is the case, then all of the definite descriptions in TF contribute the exact same thing -- Omni. Consequently, TF becomes:

(1) Phi (Bach wrote fugues).
(2) Phi (Omni=Omni).
(3) Phi (Omni=Omni).
(4) Phi (Stalin was evil).

But in that case, TF becomes patently bankrupt. For no two of the following sentences are logically equivalent: “Bach wrote fugues”, “Omni=Omni”, “Stalin was evil”. Consequently, the fact that “Phi” is PSLE doesn’t guarantee *any* of the steps in TF.

There is only one way to save TF. First of all, as we saw, if TF is to work, definite descriptions must be treated as referring expressions. Further, if TF is to work, we need the semantic contribution of a referring expression to something other than, or at least additional to, its referent. So we must say: Two co-referring expressions can make *different* semantic contributions. So, for example, “the class C such that x falls in C iff x=x” makes a *different* semantic contribution from “the class C such that x falls in C iff (x=x and Bach wrote fugues”, even though *both* those expressions co-refer – even though they both refer to Omni. In a word, if TF is to work, we need it to possible to change sentence-meaning by replacing a referring term in a sentence with a co-referring term. We need (SS).

This can be put another way. Suppose (SS) does *not* hold. In that case, (2) and (3) have precisely the same sense. In that case, step (3) – or, if you prefer, step (2) – becomes redundant. We might as well drop it. So TF

becomes:

(1) Phi (S)".
(2) Phi ([x:x=x and S]=[x:x=x])
(4) Phi (S*).

But this is obviously not a good argument any more. If TF is to go through, we *need* replacements of referring terms with co-referring terms to result in changes of sentence-meaning. We *need* (SS).

The other side of the coin is, once (SS) is granted, then TF does go through. But then we have the pernicious result that PSLE connectives are truth-functional.

We see now why it might be a good idea to scrutinize SS and *not* to accept Frege's thesis that sentences refer to truth-values. Taken by itself, that particular thesis may be plausible. But the assumptions that generate it also generate absurd conclusions – there is only proposition, only one fact, all PSLE connectives are truth-functional. Of all of those assumptions, only one is less than self-evidently correct. That one is (SS). I now propose to show why, in my view, (SS) is wrong.

Chapter 5. *why SS is wrong*

We've seen some reason to think that the Slingshot may involve a fallacy. We don't have many options. The Slingshot goes through iff all of (LL), (CR), and (SS) are correct. It is pretty clear that (LL) and (CR) are correct. So our only option is to reject (SS).

(SS) is demonstrably false. That is what I will now argue.

In this chapter, I will put forth that argument *without* discharging all of

the assumptions it involves. In subsequent chapters, I will discharge those assumptions.

If I am not mistaken, the basic conceit behind what I am about to say is found in Barwise and Perry (1990/1983[175]), though I came up with it independently. And my defense of it isn't entangled in their controversial "situational semantics" or, for that matter, in any other theoretical commitment.

Let us start with some basic points. What does it mean to say that "Socrates" refers to Socrates? It means this:

(*) Sentences of the form "…Socrates…" mean:…*Socrates*…

Put another way, it means that, when "Socrates" occurs in a sentence, *Socrates* is the thing semantically contributed by that expression to the sense of that sentence.

Consider the sentence

(a) "Socrates smokes".

Suppose that were made true by *Plato's* smoking, but that *otherwise the semantic rules of English were the same as they are now*. In that case, "Socrates" would plainly refer to *Plato* (at least in that context), not to Socrates.

(Of course, if "smokes" meant "is such that Plato smokes", then "Socrates" could refer to Socrates even though the truth of (a) depended on Plato's, not Socrates', being a smoker. But the italicized qualification in the last paragraph forestalls that kind of problem for my analysis. My point is that, if "Socrates smokes" were made true by Plato's smoking, but that *apart from*

that the rules of English semantics were just what they are now, *then* "Socrates" would refer to Plato, at least in the context of (a). So the ideas is that if "smokes" means what it actually means – namely, *smokes*, and not *is such that Plato smokes* – then "Socrates" would refer to Plato.)

Under the circumstance described, "Socrates" would not refer to Socrates in (a) because *Plato*, not Socrates, would be the semantic contribution of "Socrates" to the sense of that sentence.

So if "Socrates" is to refer to Socrates, it is *necessary* that, when "Socrates" occurs in a sentence, Socrates be the semantic contribution of that expression to the sense of that sentence.

It is also *sufficient.* Suppose somebody said the following:

I grant that sentences of the form "…Socrates…" mean:…*Socrates.*…I grant that "Socrates smokes: means that *Socrates* smokes, and that "Plato kicked Socrates" means that Plato kicked *Socrates.*

But that doesn't mean that "Socrates" *refers* to Socrates. That is, no doubt, a *symptom* of the fact that "Socrates" refers to Socrates. But that is not what the relation of reference *per se* consists in.

This is patent nonsense. If "Socrates" unfailingly contributes *Socrates* to sentences in which it occurs, then *ipso facto* "Socrates" refers to Socrates. If Socrates is the semantic value of "Socrates", then "Socrates" refers to Socrates.

There are important and interesting questions about what kinds of facts *ground* the reference relation. But these must not be confused with questions about what reference *is.*

Russell[176] once said that "fire" refers to fire because, at some

psychological level, we react to tokens of the word “fire” the way we would react to fire itself. These days, many say that “Socrates” refers to Socrates because tokens of “Socrates” have a certain kind of causal connection to Socrates. This is known as the “causal” theory of reference.[177]

These theories may or may not be correct. But they are theories about what *grounds* reference. But they are not theories about what reference itself is. There is no need for a *theory* of that. And there isn’t (or shouldn’t be) disagreement about that.

A causal theorist will admit that “Socrates” refers to Socrates just in case “…Socrates…” means…*Socrates*… I cannot imagine a causal theorist saying the following:

I grant that Socrates is the thing contributed by “Socrates” to the senses of sentences in which that expression occurs. I grant that “Socrates smokes” is true iff *Socrates* smokes. But that has nothing to do with the fact that “Socrates” *refers* to Socrates. For “Socrates” to refer to Socrates is just for some causal connection to hold between the man and tokens of the expression. It is totally irrelevant what is semantically contributed by such tokens.

That would be idiotic. The causal theorist says (or should say):

We know that “Socrates” refers to Socrates. We know that “…Socrates…” means:…*Socrates*…We know that “Socrates smokes” is true iff *Socrates* smokes. I don’t contest any of that.

What *I* want to know is this: What is relation between “Socrates” (or tokens thereof) and Socrates that *grounds* the facts I just described? In virtue of what do tokens of Socrates contribute Socrates, as opposed to Plato, to the senses

of sentences of the form "…Socrates…"?

Clearly "…Socrates…" means:…*Socrates.*…And clearly for "Socrates" to refer to Socrates just *is* for "Socrates" to contribute Socrates to the meanings borne by sentences in which that expression occurs.

But it is not a *brute* fact about the world that "…Socrates…" mean… *Socrates.*…Surely that fact can be *explained* – traced to more fundamental facts. I want to know what those more fundamental facts are.

My theory is: some kind of *causal* relation grounds that fact. The *reason* Socrates is the semantic contribute of "Socrates" is that tokens of the latter have some kind of *causal* relation to Socrates (or, at any rate, to states of affairs involving Socrates).

But the reference-relation – the relation which I am investigating and whose *grounds* I am trying to uncover – is given by the formula: x refers to y iff "…x…" mean:…*y*…

This banal point about reference is the first step in our response to Frege.

a refinement

The banal point just made needs to be qualified. The best way to see how it needs to be qualified is to consider an objection to it:

You say that for "Socrates" to refer to Socrates just *is* for "…Socrates…" to mean…*Socrates.*…

But that exact same formula "shows" that expressions like "Sherlock

Holmes" and "the round square" refer to things. Surely the following is true: Sentences of the form

"….Sherlock Holmes…" mean (or are true iff):…*Sherlock Holmes.*…

And sentences of the form

"…the round square…" mean (are true iff) iff:…*the round square.*…

But there is no Sherlock Holmes; so (at least arguably) "Sherlock Holmes" doesn't refer to anything. And there is no round square; so "the round square" couldn't possibly refer to anything. It seems, then, that our schema – "O" refers to O iff "…O…" means…*O.*…-- is too permissive. For if that schema is right, then given any expression, we can make it refer to something merely by fiddling about with quotation marks and the "means that" connective.

The objector is right to this extent: to avoid the charge of vacuousness, our formula does need a bit of refining. The needed refinement is this.

Consider an expression "E". For E to be a term that *refers* to something is for there to be some object O such that sentences of the form "…E…." mean: ….*O.*…Equivalently, for "E" to refer to O is for "…E…" to mean *O has the property…x.*…

So *if* "the round square" really refers to something, here is what that amounts to. There is some object RS such that "….the round square…" means:…*RS.*…

There is no object RS having that property. So "the round square" doesn't manage to refer to anything.

What we've just said is neutral between the view that definite descriptions are referring terms (i.e. are terms that, given their syntactical roles, are *supposed* to refer) and those who hold, with Russell, that such terms are quantifiers are certain kind.

According to the theory of descriptions, "the round square" doesn't refer to anything because it isn't the kind of thing that is even *supposed* to refer to anything. According to that view, "…the phi…" means: *exactly one thing has phi and anything having phi also has the property…x…*So "the inventor of bifocals smoked" means: *exactly one thing invented bifocals and anything having that property smoked.* So there is *no* object O such that "the inventor of bifocals smoked" means: *O smoked.* Of course, everybody agrees that if "the inventor of bifocals smoked" is to be *true*, then there had to *be* some individual (say, Charlie) who invented bifocals who smoked. But, if Russell's theory is right, "the inventor of bifocals smoked" doesn't mean: Charlie smoked. According to Russell's theory, "the inventor of bifocals smoked" expresses the existence claim mentioned a moment ago. So if the theory of descriptions is right, the "the phi" never refers, not even if there is a unique phi.

But, given only that there is no object O such that "…the round square…" means:…*O…*, it doesn't follow that the theory of descriptions is correct. The referentialist (somebody who believes that definite descriptions are referring expressions) doesn't say that *all* definite descriptions actually manage to pick out objects.

The referentialist says this: if there is[178] some object O that uniquely has phi, *then* "…the phi…" means:….*O*….But if there is no such object, then "….the phi…" is abortive – just as a token of "that is lovely", where "that" didn't pick anything out (because the speaker is hallucinating), would be abortive.

So the referentialist says: *if* there is a unique object RS that is both round and square, then "….the round square…" means: *RS has…x…* And the referentialist says that if there is no such object, then "…the round square…" does not (at the level of literal meaning) encode any proposition.

Obviously there is no such object. Russellians and referentialist agree on that. So they agree on whether there is something to which "the round square" refers. The difference is that, according to Russell, "the round square" isn't even *supposed* to refer to anything.

Our analysis of reference is consistent with the fact that "the round square" doesn't refer to anything, and it is neutral on the question whether definite descriptions are referring terms (i.e. are terms that are *supposed* to refer). Our analysis doesn't say that "the inventor of bifocals" *does* refer. Our analysis says only that *if* that expression referred to (say) Charlie, *then* "….the inventor of bifocals…" would mean that Charlie has…x…

This enables us to deal with the Sherlock Holmes point. There is no Sherlock Holmes (or so I believe). But *if* "Sherlock Holmes" actually referred to some something, that would amount to there would be some object *SH* such that "…Sherlock Holmes…" means: *SH has…x…*(or…*SH…*).

Of course, some people say that "Sherlock Holmes" *does* refer – to a fictitious entity or some such thing. Very well. Let *SH* be that entity. In that case, according to my analysis, "Sherlock Holmes" refers to SH in virtue of the fact that "…Sherlock Holmes…" means:…*SH…*

Now "Socrates" actually does refer to something; it refers to Socrates. According to my analysis, that consists in the fact that "…Socrates…" mean: …*Socrates…*In general, for an expression to *refer* to a thing just *is* for that thing to be the semantic contribution of that expression. This does not mean that the causal theory is right or wrong, or that any other specific "theory" of reference is right or wrong. The causal theorist surely grants that "Socrates"

semantically contributes Socrates to “Socrates was wise”; and he also grants that if “Socrates” contributed the planet Neptune to that context, then “Socrates” would *not* refer Socrates (at least not in that context). The causal theorist, it seems to me, grants that “Socrates” semantically contributes Socrates; and he also grants that this is what it is for “Socrates” to refer to Socrates. The causal theorist says that *this* fact – the fact that “Socrates” contributes, and thus refers to, Socrates – is grounded in, and possibly even identical with, some fact about the causal relation between tokens of “Socrates” and Socrates. But, it seems to me, what the causal theorist is doing is not so much to deny that for “Socrates” to refer to Socrates is for that expression to semantically contribute that entity; rather, the causal theorist is saying “that expressions’ contributing/referring to that entity is identical with, or supervenient on, that expressions having a certain causal relation to Socrates”.

The basic point is this: for “E” to refer to an object O just is for sentences of the form “…E…” to mean:…*O*…(or *O has…x*…) “Plato” refers to Plato in so far as, and only in so far as, sentences of the form “…Plato…” mean: *Plato has…x*…

Summary

It seems reasonable to say that for an expression E to refer to O is simply for sentences of the form “…E…” to mean:…*O*…

That thesis constitutes the first step in our attack on Frege’s slingshot.

So far we’ve given some prima facie reasons for this thesis. Later we will give further grounds.

the referential-attributive distinction

Here is the next step in our attack on Frege's slingshot. Consider the list of things that grammarians – not logicians or philosophers – refer to as "nouns" and "noun phrases": "Socrates", "Plato", "somebody", "nobody", "the tallest man alive", "three birds".

Everything on this list is *either* a referring term ("Plato") *or* a quantifier ("somebody").

Definite descriptions are on this list. It is an open question whether, at the level of semantics (literal meaning) definite descriptions are referring terms or quantifiers.

Intuitively, it seems that "the even prime" is a referring term that refers to the even prime (the number two).

Russell compellingly argued that "the even prime" is a certain kind of quantifier.

It is a matter of controversy which view is the right one.

What is *not* a matter of controversy is this. Definite descriptions *function* (pragmatically if not semantically) both as referring expressions *and* as quantifiers.[179] In some cases,

(*) "the man two feet to your left smokes"

communicates a proposition of the form

(**) *so and so* smokes.

In such a case, "the man in the corner" is *functioning* as a referring term: it

picks out so and so.

In other cases, (*) communicates a proposition of the form:

(***) *There is exactly one man two feet to your left and any such man smokes.*

Here "the man two feet to your left" is functioning as a quantifier, as what linguists call a "generalized quantifier".

In the first case, the definite description is functioning "referentially". (It is functioning as a referring term.) In the second, the definite description is functioning "attributively", as a kind of quantifier.

It is open what the correct *semantic* analysis of "…the phi…" is. In other words, it is open what the *literal meaning* of "the phi". Russell says it its a quantifier; Strawson thinks it is a referring term; Barwise and Perry think it is semantically ambiguous.

What everybody agrees on is this: "…the phi…" is capable of *functioning* both referentially and attributively. (Even Strawson admits this.[180]) For our purposes, that is all we need to know.

Now we can evaluate Frege's argument. First of all, *if* both CR and SS are correct, *then* Frege is surely right: all true sentences do stand for the same thing, and that thing is the property of being true. Frege's argument is valid.

CR is a correct principle.

This means that SS must be false.

And, demonstrably, SS *is* false. Let us now turn to this issue.

the falsity of SS

I will now try to show that SS is false. Before giving my argument in full, I

would like to give a rough statement of its overall structure.

the argument against SS: an outline

Consider the following two sentences:

(i) The greatest hemlock-drinking philosopher of antiquity antagonized many Athenians.

(ii) The man who figures as the main protagonist in most of Plato's dialogues antagonized many Athenians.

(ii) is what results when a referring term in (i) is replaced with co-referring term. And yet (ii) has a different sense – bears a different proposition from (i).

Wrong!

Consider (i). We can read it either referentially *or* attributively. Let us consider each case.

Suppose we read it attributively. In that case, "the greatest hemlock-drinking philosopher of antiquity" isn't a referring term at all. If that expression isn't functioning as a referring term, then (ii) is not the result of replacing a *referring* term (i) with a co-referring term.

(ii) results from (i) by replacing "the greatest hemlock-drinking philosopher of antiquity" with "the man who figures as the main protagonist in most of Plato's dialogues".

If "the greatest hemlock-drinking philosopher of antiquity" is operating attributively, then it is not functioning as a *referring term.* And in that case, (ii) is not what results when a *referring* term in (i) is replaced with a co-

referring term.

There is absolutely no doubt that (i) and (ii) *communicate* different propositions. (i) communicates (inter alia) that anyone having the property of being a greatest philosopher of antiquity also had the property of antagonizing many Athenians.

(ii) doesn't communicate that at all.

And (ii) communicates (inter alia) that anyone having the property of being the main protagonist in most of Plato's dialogues also has the property of antagonizing many Athenians.

(i) doesn't communicate that at all.

So (i) and (ii) communicate different things. Nobody denies this. (As for whether the propositions they *semantically* encode are different, that is an open question. But it is clear that – pragmatically, if not semantically -- they *communicate* different propositions.)

But (i) and (ii) communicate different propositions only in so far as they are read attributively. In other words, (i) and (ii) communicate different propositions – have different senses -- only in so far as the definite descriptions are operating *not* as referring terms, but as quantifiers.

So in so far as (i) and (ii) differ in sense, that is because the definite descriptions are, in effect, *not* referring terms.

But in so far as they are not referring terms, CR does *not* guarantee *anything* about (i) and (ii) having the same referent. CR says that reference is preserved when referring terms are replaced with co-referring terms: it doesn't say anything about what happens when *quantifiers* are replaced with other quantifiers.

Let us extend the point just made. Consider the following sentences:

(iii) Plato snored.

(iv) The unique thing x such that x is identical with Plato and arithmetic is incomplete snored.

Here is the *obvious*, but wrong, thing to say about these sentences.

(iv) is what results when the referring term in (iii) is replaced with a co-referring term. Thus, by CR, (iii) and (iv) must refer to the same thing, if they refer to anything.

At the same time, (iii) and (iv) plainly differ in sense. (iv) says (inter alia) that arithmetic is incomplete. (iii) does not say this.

Here is the *correct* thing to say about these sentences. There is certainly *a* legitimate reading of (iv) whereby it has a different sense from (iii). That reading of (iv) is the attributive reading. It is the reading on which the definite description is a quantifier.

On that reading, (iv) does not result from (iii) when a referring term in (iii) is replaced with a co-referring term. On that reading, (iv) is what results when a referring term in (iii) is replaced with a *non*-referring term – with a quantifier.

There is another way to make this point. Frege thinks that one can produce a change of sentence-*sense* by replacing referring terms with referring terms.

(v) The inventor of bifocals smoked.

(vi) The first post-master general smoked.

(vi) has a different sense from (v); and (vi) is what results from (v) when a referring term in the latter is replaced with a co-referring term.

But suppose that we treat the definite description in (v) as a *referring term.* In that case, its semantic contribution to the sense of (v) is the thing it refers to. (This is a truism.) The thing it refers to is Benjamin Franklin.

Thus, if we treat "the inventor of bifocals" as a *referring* term, then the proposition borne by (v) is:

(v_R) Benjamin Franklin smoked.

I am not saying that, in fact, "the inventor of bifocals" is a referring term. [181] That would be a controversial point. I am saying that *if* "the inventor of bifocals" is functioning as a referring term in (v), then its semantic contribution to the proposition borne by (v) is the thing it refers to. That is *not* a controversial point.

That thing is Benjamin Franklin. Thus, *if* "the inventor of bifocals" is functioning as a referring term in (v), *then* the proposition borne by (v) is (v_R).

Suppose that, in (vi), "The first post-master general" is functioning as a referring term. In that case, its semantic contribution to the proposition borne by (vi) is the thing it refers to: Benjamin Franklin. And in *that* case, the proposition borne by (vi) is simply (v_R).

If the definite descriptions in (v) and (vi) are functioning as *referring terms*, then both those sentences bear the very same proposition.

So in so far as those sentences bear *different* propositions, it is entirely because the definite descriptions are not functioning as referring terms. In so far as those sentences bear *different* propositions, it is because those sentences are being read attributively: the definite descriptions are being

treated as *quantifiers*, not referring terms.

Obviously there is a natural reading of (v) and (vi) whereby they have different propositions. But on that reading, it is *not* the case that (vi) results from (v) by replacing a referring term with a co-referring term. A *quantifier* has been replaced with another quantifier.

Church's slingshot

Let us now consider Church's argument.[182] The core of that argument is given by the following sequence of sentences.

(i) Sir Walter Scott is Sir Walter Scott.
(ii) Sir Walter Scott is the author of Waverly.
(iii) Sir Walter Scott is the man who wrote twenty-nine Waverly novels altogether.
(iv) The number n, such that Sir Walter Scott wrote n Waverly novels, is twenty-nine.
(v) The number of counties in Utah is twenty-nine.

By CR, (i) and (ii) and (iii) must all refer to the same thing.
(iii) and (iv) are, in Church's view, synonymous (or "very nearly so").
Synonymous expressions must co-refer. So (iii) and (iv) co-refer.
By CR, (v) refers to the same thing as (iv).

It follows that (i) and (v) refer to the same thing. CR guarantees that reference has been preserved in the journey from (i) to (v).

The *only* semantically significant thing which (i) and (v) have in common is the truth-value.

So reference has been preserved, and nothing *besides* truth-value has been

preserved.

So (i) and (v) co-refer – they refer to the truth-value *true.*

What is true of (i) and (v) is true (mutatis mutandis) of *any* two true sentences. Let S and S* be any two true sentences you choose. Given CR, plus the innocuous principle that synonymous expressions co-refer, a little bit of syntax-chopping shows that S and S* have the same referent.

Thus any two true sentences refer to the same thing: the property of truth.

An exactly similar argument shows that any two false sentences refer to the same thing: the property of falsity.

evaluating Church's slingshot (an outline)

Let us evaluate this argument, beginning with the first two steps:

(i) Sir Walter Scott is Sir Walter Scott.
(ii) Sir Walter Scott is the author of Waverly.

These sentences are the same except that, where the first contains "Sir Walter Scott", the second contains "the author of Waverly". CR guarantees that (i) and (ii) have the same referent *only if* "the author of Waverly" is a referring term. CR doesn't say anything about what happens when terms are replaced with things *other* than co-referring terms.

Suppose we read "the author of Waverly" as a referring term. In that case, the semantic contribution of "the author of Waverly" is simply the thing it refers to. That thing is Sir Walter Scott. So, read referentially, the proposition borne by (ii) is:

(ii_R) Sir Walter Scott is Sir Walter Scott.

Of course, the proposition borne by (i) is (ii_R). So if "the author of Waverly" is functioning as a referring term, there is *no* difference in sense between (i) and (ii).

Exactly similar reasoning shows that, if (iii) is read referentially, it bears (ii_R).

If the definite descriptions in (i)-(iii) are read *referentially*, then there is *no* difference in sense among those propositions. In that case, of course, Church's argument is blocked at the outset.

So if Church's argument is to work, then those definite descriptions must be read attributively, i.e. as quantifiers. But then the transition from (i) to (ii) isn't licensed by CR. Neither are the transitions from (ii) to (iii) or from (iv) to (v).

Chapter 6. the argument proper

We've just given an *outline* of an argument against Frege's slingshot and also against Church's slingshot.

That outline falls short in two ways. First, it involved two controversial (but, I believe) correct assumptions. So I must discharge those assumptions if the argument just outline is to be cogent.

Those assumptions are:

(AT) When a referring term is functioning attributively, it is functioning as a quantifier.

(SC) For E to refer to O simply *is* for sentences of the form "…E…" to bear

the proposition:…*O*…

Obviously those two assumptions must be shown correct if the argument just outlined is to work.

But that argument is inadequate in another way. It involved an assumption that is, I think, false. However, the falsity is innocuous because, in that argument, that assumption was really just functioning as a compressed version an entirely correct assumption - albeit one that couldn't have been articulated within the framework of a summary.

That (arguably) false assumption is this:

(QRT) Quantifiers are not referring terms.

Many (including myself) hold that quantifiers *are* referring terms (of a very special kind).

Obviously if the argument just outlined is to work, then I must show that its reliance on QRT is totally innocuous and that QRT really was just functioning in it as a device of abbreviation.

SC. Then we will show why our argument doesn't really rely on QRT.

In this chapter I will argue for SC and AT.

In chapter 7, I will show why my argument doesn't involve QRT (except as a kind of typographical short-hand).

what is semantically contributed by referring terms? my argument for (SC)

According to Frege[183], Mill[184], Church[185], and others "the phi" is a referring term that *also* semantically contributes its sense.

This is a natural position to take. Nothing seems more obvious than that in

“the inventor of bifocals was a many of many talents”, the definite description contributes not only a certain individual (Benjamin Franklin) but also some descriptive content.

Before we can assess this, we must distinguish very clearly between what is *semantically* contributed by an expression, on the one hand, from what is *cognitively* contributed by it, on the other. What an expression conveys is not always confined to its literal meaning (its semantic content). There can be no doubt “the inventor of bifocals”, when it occurs in a sentence, *cognitively* contributes a certain descriptive content; such content is communicated or pragmatically imparted. But not everything that is conveyed (cognitively contributed) by an utterance or expression is literally meant by it. And as I will try to show, there can be no doubt that, if “the phi” really is a referring term, then its sole *semantic* contribution is a certain individual.

For the sake of argument, let us suppose that “the inventor of bifocals” is a referring term, and investigate the consequences of that assumption. The most obvious and immediate consequence is that, for some object O, sentences of the form “…the inventor of bifocals…” are made true by O’s having…x…, and such sentences are false of O does *not* have that property.

Let us suppose (what is probably true) that Benjamin was a unique inventor of bifocals. In that case, the proposition encoded in “the inventor of bifocals snored” is true just in case Benjamin Franklin snored. Suppose that the proposition encoded in that sentence could be true regardless of whether Benjamin Franklin snored. In that case, it is very hard to see how “the inventor of bifocals” could be said to be a term that referred to Benjamin Franklin. Presumably, if “the inventor of bifocals” refers to Benjamin Franklin, then the proposition encoded in “the inventor of bifocals snored” should be *about* Benjamin Franklin. And surely that proposition couldn’t be about Benjamin Franklin unless its truth turned on whether Benjamin

Franklin snored.

Russell himself thought that the proposition encoded in “the inventor of bifocals snored” didn’t really say anything about the person who in fact invented bifocals – he said that that proposition was:

(BF) *exactly one entity invented bifocals and any such entity was a snorer.*

Of course, (BF) could be true in worlds where there was no Benjamin Franklin. In fact, given any individual, (BF) could be true in a world lacking that individual. (Russell himself doesn’t use possible-worlds talk. But, in this context, such talk helps to illustrate his commitments.) So if Russell is right, then “the inventor of bifocals” doesn’t refer to any body. The essence of reference seems to be this: E refers to O iff “…E….” encodes a proposition to the effect that…O…

So *if* “the inventor of bifocals” is a referring term – and I am leaving it open whether it is – *then* surely Benjamin Franklin (or whoever it was that invented bifocals) is what is semantically contributed by that expression.

The Frege-Mill position is that “the inventor of bifocals” *also* contributes some kind of descriptive information. So the proposition encoded in “…the inventor of bifocals…” is *not* merely “…Benjamin Franklin…”

Here we must be careful. Given that Frege and Mill hold that “the inventor of bifocals” is a referring term, they are committed to holding (though they may not in fact hold) that *part of* what is said by “the inventor of bifocals snored” is that Benjamin Franklin snored. This much is not up for grabs. If they make the proposition encoded in “the inventor of bifocals snored” be true in worlds where there is no Benjamin Franklin, then (whether they knew it or not), they are, by implication, *denying* that “the inventor of bifocals” is a referring term.

So Frege-Mill are committed to the following view. *Part* of what is said by "the inventor of bifocals snored" is that Benjamin Franklin snored. (More accurately, if Frege and Mill are right, then for some object O, part of what is said by "the inventor of bifocals snored" is: *O snored*.) But, arguably, they also hold that "the inventor of bifocals" semantically contributes some kind of descriptive content.[186]

The question is: *How* is this descriptive content contributed? In other words, *what* is the proposition that, if Mill and Frege are right, is encoded in "the inventor of bifocals snored"? We've seen that, for them, it is not merely: *Benjamin Franklin snored*. But if, as they maintain, "the inventor of bifocals" is a referring term, *Benjamin Franklin snored* must be a *part* of what is encoded in "the inventor of bifocals snored".

I don't see how their view can made coherent. Surely they are not saying that

(IBS) "the inventor of bifocals snored"

means

(i) *Benjamin Franklin snored; and somebody uniquely invented bifocals and any such person snored.*

It would be extremely unnatural to suppose that it is the proposition encoded in (IBS). Not all sentences of the form "…the phi…" express *conjunctive* propositions. Frankly, I don't think it is really worth arguing against the idea that (i) gives the semantics for (IBS). The idea is too tortured to consider. Indeed, no one has ever proposed it.

So we are left with the question: *How* is the descriptive content associated

with “the inventor of bifocals” to make it into the proposition encoded in (IBS)? (Remember that if, as Mill and Frege maintain, “the inventor of bifocals” is a referring expression, then at least *part* of what (IBS) says is: *Benjamin Franklin snored*.)

Another option is to say that (IBS) means:

(ii) Benjamin Franklin snored and Benjamin Franklin invented bifocals.

But, like (i), it is simply too unnatural to suppose that (ii) gives the meaning of (IBS). And no one has ever proposed it.

the issue between Russellians and Strawsonians

We get some leverage on this issue the moment we make it clear what the difference is between those who hold that definite descriptions are referring terms and those who hold that they are quantifiers. Let us describe the latter as “Russellians” and the former as “referentialists”.

The difference between referentialists and Russellians is subtler than is usually thought. The only difference is that, in the semantic rule for “the phi”, the referentialist puts the “means that” operator a few inches to the right of where the Russellian puts it.

According to the Russell, the semantic rule for “the inventor of bifocals” – the rule that assigns it meaning – is this:

(RUS) Sentences of the form “…the inventor of bifocals…” mean that:

exactly one individual O invented bifocals and…O…

According to the Referentialist, the semantic rule for “the inventor of bifocals” is this:

(REF) If there is exactly one individual O such that O uniquely invented bifocals, then sentences of the form

“…the inventor of bifocals…”

mean that:

…*O*…

And if there is *no* unique individual who invented bifocals, then

“…the inventor of bifocals…”

doesn’t encode any proposition.

In general, according to the Russellian, the semantic rule for “the phi” is:

“…the phi…”

means that:

exactly one individual O has phi and…O….

According to the referentialist, the semantic rule for “the phi” is:

if there is a unique individual O having phi, then

“….the phi…”

means that:

…O…

If there is no individual uniquely having phi, then

“…the phi…”

doesn’t have any proposition for its literal meaning.

(RUS) and (REF) are very similar. Speaking loosely, the only difference lies in where the “means that” operator is placed. (REF) gives it *narrow* scope. (RUS) gives it wide scope. Apart form that, (RUS) and (REF) are in (near) perfect agreement.

But this difference in scope is important. An immediate consequence of it is that, for advocates of (REF), “…the phi…” are *de re* about the unique individual (if any) having phi, whereas for advocates of (RUS), “…the phi…” is only de dicto about that individual.

Both (REF) and (RUS) agree that “the phi” is semantically associated with descriptive information. In fact, both (REF) and (RUS) agree that “…the phi…” will be true if and only if there is exactly one object O having phi

and…O…

Advocates of (REF) couldn't agree more that "the phi" is semantically associated with descriptive content; that, as a matter of semantics, in order for "…the phi…" to be true, it is necessary and sufficient that there be exactly one object O having phi and that O have…x…

The difference between (REF) and (RUS) is that, for (REF), this descriptive content has a reference-*fixing*, not a meaning-*giving* role.

Here, of course, I am alluding to Kripke's (1972) important distinction. If I say "Gödel was the guy who proved the incompleteness of arithmetic", I am telling you who "Gödel" refers to; I am not saying that the *meaning* of "Gödel" includes anything having to do with mathematical proofs. More accurately, I am saying:

(*) Somebody x uniquely proved the incompleteness of arithmetic and "Gödel" refers to x.

I am *not* saying:

(**) "Gödel" is synonymous with: *the guy who proved the incompleteness of arithmetic.*

(REF) says: "the inventor of bifocals" is a referring term. As a matter of semantics, it (or tokens of it) pick out that individual who uniquely invented bifocals, and it picks out nothing if there was no such individual. So for (REF) the descriptive content associated with "the inventor of bifocals" serves to pick out *to whom* "the inventor of bifocals snored" attributes the property of snoring. But that descriptive content does not *itself* figure in the proposition encoded in that sentence.

(RUS) says: that descriptive content *does* figure in that sentence. That sentence says of that descriptive content that it is uniquely instantiated and that anything instantiating it snores.

But notice that, if (REF) is right – I leave it open whether it is – then anyone who has the foggiest notion of the semantics of "the inventor of bifocals" know that "the inventor of bifocals" will certainly not be *true* if there is no inventor of bifocals[187], and also knows that it will not be true if there are a million (equally contextually salient) inventors of bifocals. So if (REF) is right, then when you say to me[188]:

(IBS) "the inventor of bifocals snored"

I know that, if your utterance is to be (literally) true, the following must be the case:

(*) somebody uniquely invented bifocals and any such individual snored.

So, if (REF) is right, then (IBS) will *communicate* (though it won't semantically encode) that a certain description is uniquely satisfied. So descriptive content will certainly be *communicated.*

So the supposition that (REF) is correct is compatible with several important facts. First, (IBS) is true only somebody uniquely invented bifocals. Second, (IBS) communicates (though it does not, if (REF) is right, semantically encode) the proposition: *somebody uniquely invented bifocals.* Third, the association between the expression "the inventor of bifocals" and the corresponding descriptive information is *semantic*: as a matter of semantics, "the inventor of bifocals" refers (if at all) to that individual who

uniquely invented bifocals.

Let us now return to Frege and Mill. Frege and Mill want to satisfy three desiderata:

(i) They (rightly) want to do justice to the fact (ISB) certainly *communicates* something about bifocal inventing.

(ii) They (rightly) want to do justice to the fact that there is a *semantic* – not a chance or purely circumstantial – connection between the description *inventor of bifocals* and the expression "the inventor of bifocal".

(iii) They want to say that "the inventor of bifocals" is a *referring term*. (This is a controversial, but reasonable, desideratum.)

There is a very natural way to accommodate all these desiderata. It is take the position that the semantics of "the inventor of bifocals" is given by (REF). As we've seen, (REF) accommodates all the desiderata Frege and Mill want to be accommodated.

The other side of the coin is that if, like Frege and Mill, you believe that definite descriptions are referring terms, then it becomes extremely difficult – impossible, frankly – to accommodate (i) and (ii) without taking a position more or less like (REF).

Now we can close the argument. *If* definite descriptions are referring terms, then given that (REF) is (at least approximately) correct, it follows that the proposition semantically encoded in (ISB) is:

Benjamin Franklin snored.

The descriptive content associated with “the inventor of bifocals” doesn’t make it into the proposition semantically encoded in (tokens of) (ISB).

We are not 9/10th’s of the way towards vindicating our thesis that,

(R) if E is a referring term whose referent is O, then O is the *sole* semantic contribution of occurrences of E.

There is only one small bridge to cross before we can establish (R). Definite descriptions are not the only referring terms. (In fact, some[189] say they aren’t referring terms at all.) So far we haven’t established that the sole semantic contribution of “Socrates” is Socrates – we haven’t ruled out that it contributes some kind of descriptive content. And this is just what Frege thought.

There are two reasons why this doesn’t threaten our claim. First, Kripke famously showed that “Socrates” is not (semantically[190]) associated with any description (or, more exactly, with any uniquely individuating description). What Kripke said makes it very hard to believe that “Socrates” semantically contributes anything other than Socrates.

The other reason is this. For the sake of argument, suppose that Kripke is wrong and that “Socrates” does, at the level of semantics, have a sense. What exactly is a sense? It very much seems to be the kind of thing that is given by a description. Now if the sense of “Socrates” is a description, the description in question must be one that is supposed to *single something out*. That description will obviously not be something like “is a man”, which applies to many things. It must be something that, supposedly, singles out a single object. Thus, if the “Socrates” is given by a description, it must a *definite*

description.

So if Kripke is right, then it becomes hard to see how “Socrates” could fail to be synonymous with some definite description (or perhaps with different descriptions). In that case, we simply use the arguments just developed to show that “Socrates – *qua* expression synonymous with (say) “the greatest hemlock drinking philosopher of antiquity” – contributes, if anything, some individual, and nothing besides.

I should deal with a worry:

The expression “the teacher of Plato” refers to the teacher of Socrates. (In any case, this is a reasonable and widely held view.) But no one thinks that the “the teacher of Plato” contributes the teacher of Plato to the sense of a sentence. This seems to devastate your claim that for E to refer to O is for tokens of E to contribute O to the senses of the sentences in which those tokens occur.

Let us suppose that, as the objector maintains, “the teacher of Plato” is a referring term. In that case, it refers to Plato. In that case, when “the teacher of Plato” occurs in a sentence, then Socrates is in fact the thing semantically contributed by that occurrence. A consequence is that sentences of the form “…the teacher of Plato…” bears a proposition of the form:...*Socrates*.…

Of course, there are compelling reasons to *deny* that “the teacher of Plato” contributes Socrates to sentences in which it occurs. Russell denied just this. But this denial is nothing other than a denial of the view that “the teacher of Plato” is a term that refers to Socrates.

The objector has a legitimate worry; and his hostility towards my analysis

of reference is an understandable expression of that worry. But it is not the right way of dealing with that worry.

The worry is as follows. First of all, it is very *not* to believe that "the teacher of Plato" is a referring expression. Russell and Frege have warned us not to read too much into grammatical surface-structure. But somehow our intuition that "the teacher of Socrates" is a referring term lives on.

At the same time, in at least *some* legitimate sense of the word "understand", one can plainly understand a sentence like "the teacher of Plato was wise", and "the richest man in Holland sleeps in a sensory deprivation tank" without having any idea at all who taught Plato and who has more money than anyone else in Holland. Now if, indeed, "the teacher of Plato" contributes Socrates to sentences in which it occurs, then presumably I couldn't *understand* sentences like "the teacher of Plato was wise" without knowing that *Socrates* specifically was the fellow to whom wisdom was being attributed. And yet, as we've seen, there is plainly some sense in which I can understand the sentence "the teacher of Plato was wise" *without* knowing that *Socrates* is involved in any way.

Russell dealt with the *second* of these two points by denying that "the teacher of Plato" is a referring term (or, at least, that it is a term that refers to individuals). But not everyone is willing to follow Russell here. Most want to hold onto the idea that "the teacher of Plato" is what, grammatically, it seems to be: an expression that refers to some ancient Greek person. *But* they want to grant that one can understand "the teacher of Plato was wise" without knowing that *Socrates* must be wise for that statement to be true. The resulting conception is this: "the teacher of Plato" is a referring expression *but it doesn't semantically contribute Socrates* (even though it obviously refers to Socrates, given that Socrates was, in fact, the teacher of Plato).

For the sake of clarity, let me re-affirm the position I am advocating. If

"the teacher of Plato" *does* refer to Socrates, then "…the teacher of Plato…" *does* encode the proposition "…Socrates…"

How do I deal with the problem just described? The key is to realize that literal meaning and cognitive significance – semantics and epistemology – often pull apart quite dramatically. Clear examples of this have been produced by modern semanticists. Suppose we decide to add the following semantic rule to English: If the word "Argo" occurs in a sentence S, that occurrence semantically contributes that person, whoever it is, who is the person currently alive who will be the foreign minister of Zimbabwe in exactly twelve years. So occurrences of "Argo" contribute a *person*, not a concept or description. You and I certainly do not know who "Argo" names. (Perhaps some Zimbabweans know. But I doubt it; politics is too volatile so sustain the needed prediction.) Question: Do you and I understand "Argo is over 6 ft tall" or not?

There is a sense in which you do and a sense in which you don't. What you know is this: if there is exactly one person O such that O alone will be foreign minister of Zimbabwe in over 12 years, then, and only then, will "Argo is over 6ft tall" semantically encode the proposition: *O is over 6 ft.*

Of course, you and I don't grasp the proposition *O is over 6ft* (here I am supposing that O is the person semantically contributed by occurrences of "Argo"). (More exactly, we do not know that *that* very proposition is the one semantically encoded in "Argo is over 6 ft". There is a chance that we have met O and judged him to be over 6ft. But we surely did not know that we were therein meeting the person who, in exactly 12 years would be foreign minister of Zimbabwe. So even if you and I do grasp the proposition *O is over 6 ft,* we don't *know* that it is the proposition semantically encoded in "Argo is over 6 ft".) But we know a great deal (not everything) about what must be the case if that proposition is to be true. We don't know what it

semantically encodes; but we know a lot *about* what it semantically encodes. Perhaps it would correct to say: we are not *acquainted* with that proposition, but we know it "by description". (This is not *quite* accurate. You might *in fact* be acquainted with that proposition; but unless you know that Socrates taught Plato, you will not be able to *employ* your acquaintance with that proposition in your attempt to interpret "the teacher of Plato was wise"; you will be forced to rely on your knowledge-by-description of that proposition.) [191]

Suppose that "the teacher of Plato" refers to Socrates, but you do not know who taught Socrates. Under that circumstance, there is still *a* sense in which you can understand "the teacher of Plato was wise". It is precisely the sense in which you and I can understand "Argo is over 6ft tall". Given what we've said in connection with "Argo", and also earlier in connection the referentialist view of definite descriptions, it is clear why "the teacher of Plato was wise" might be replete with cognitive significance for somebody who didn't know that Socrates was Plato's teacher.

To sum up, if you don't know who taught Plato, then you cannot have knowledge by *acquaintance* with the proposition encoded in "the teacher of Plato was wise"). But you will have very considerable knowledge by description of that proposition.

(More exactly, even if you don't know who taught Plato, you *may* still have knowledge by acquaintance with that proposition. But if you don't know that Socrates taught Plato, you will not be able to exploit that knowledge in your attempt to interpret "the teacher of Plato was wise"; instead, you will be forced to rely on your knowledge-by-description of that proposition.)

Your having this knowledge by description is quite enough to account for the cognitive significance had for you by "the teacher of Plato was wise".

So there is no reason – nothing in semantics or epistemology – that requires us to jettison the view that "the teacher of Plato" semantically contributes Socrates (or whoever it was that taught Plato) to sentences in which it occurs. This supposition is quite consistent with the facts about what is communicated by "the teacher of Plato was wise" to someone who doesn't know that Socrates (or Gilgamesh or whoever) taught Plato. The key is remember that literal meaning sometimes outstrips knowledge by acquaintance; and that what we lack as far as a grip of literal meaning is concerned is, to some extent, compensated for by a grasp of knowledge-by-description of the literally meant proposition.

an anticipatory note

I'd like to anticipate a point we will eventually discuss at length.

It is, of course, extremely natural to believe that "the inventor of bifocals" is a referring term *and* that it contributes a sense.

It is also very hard to reconcile these two beliefs, as we saw. Russell chose the counter-intuitive path of deny that it is a referring term.

Everything falls into place when we distinguish tokens from types. For the sake of argument, suppose that Russell is wrong about "the inventor of bifocals".

There are worlds where "the inventor of bifocals" has the same semantics but refers to different people.

But surely the referent of a referring term is part of its semantic content.

If that is right, it is self-contradictory to say that "the inventor of bifocals" could have the same semantics but differ in reference.

One can take the heroic measure of saying that reference is not semantic content, even in part. But that is not a happy position.

In all worlds where English is spoken, the *type* "the inventor of bifocals" has the same semantics. But in some of the worlds *tokens* of that expression have a different semantic content; they refer to somebody other than Jefferson.

The semantic content of a *token* of "the inventor of bifocals" is the person (if any) who, in that world, uniquely invented bifocals.

The semantic content of the *type* "the inventor of bifocals" is a function that assigns an object O to tokens of that type iff O has the property of being, in the world of tokening, a unique inventor of bifocals. So the *type* has descriptive content.

But there is no *one* thing has both sense (descriptive content) and reference. The contribution of the *type* is a description. But the type doesn't refer at all. The token refers; and its referent is its only contribution.

So "the inventor of bifocals" *does* contribute both sense and reference. But not in the way that Frege thought. The *type* contributes sense; the *token* contributes reference. The proposition meant by any token doesn't include that sense. But one has to use that sense to assign the right referent to a token of "the inventor of bifocals"; one can understand that token only if one knows that to be picked out by it, an entity must satisfy a certain description. So one accesses semantic content of *tokens* of "the inventor of bifocals" through the information encoded in the associated *type.* Thus "the inventor of bifocals" ends up being expressive, even though (arguably) tokens of it are, at the level of semantics, purely denotative.

But the important point in this context – we will substantiate it later – is this: the idea that some expressions have both sense and reference is a byproduct of a failure to appreciate the type-token distinction. Types have sense; tokens have reference

the causal theory revisited

I have repeatedly said that

(SC) E refers to O iff E semantically contributes O.

But (SC) is in apparent contradiction with a very popular theory: the so-called "causal theory" of reference. That theory is:

(CTR) An expression E refers to an object O iff tokens of E have a certain kind of causal relation to O (or states of affairs involving O).[192]

I would suggest that (CTR) is *not* in fact opposed to (SC), at least not in so far as (CTR) is a coherent doctrine.

A couple of qualifications are in order. First of all, I don't think that anyone would maintain that (CTR) is meant to apply to *all* referring terms. Consider (this is a doubly tendentious example) "the first animal born in the year 3000 A.D." If definite descriptions are referring terms, as they may well be, then this is a referring term (i.e. given its syntactic role, it is *supposed* to refer to something). Suppose that Alphie is the animal satisfying that description. In that case, Alphie is the thing referred to by that expression. But "the first animal born in the year 3000 A.D." refers to Alphie by virtue of

encoding descriptive information that singles out Alphie; that expression does *not* refer to him by virtue of there being some causal relation between tokens of it and states of affairs involving Alphie. First, such a causal connection doesn't exist. And even if it did, they wouldn't be what makes "the first animal born in the year 3000 A.D." refer to Alphie. What makes "the first animal born in the year 3000 A.D" refer to Alphie is that Alphie is (or will be) the thing that (uniquely) satisfies the description encoded in that expression. So (CTR) is probably not meant to apply to definite descriptions.

Nor, for similar reasons, is (CTR) meant to apply to expressions that have constituent structure. "That snow is white" refers to the proposition *that snow is white.* The former refers to the latter because the latter is the thing having the properties in the descriptive content encoded in the former. CTR may be meant to apply to the constituents of "that snow is white", but not to that expression as a whole.

So CTR is probably not meant to apply to terms referring to propositions. [193] This may just be a repetition of the fact that CTR doesn't apply, and probably isn't meant to apply, definite descriptions; for "that snow is white" is probably synonymous with some definite description, such as "the proposition that snow is white".

(CTR) is meant, presumably, to apply to certain rigid designators – terms like "Nixon", "Socrates", and the like. (For reasons just stated, it is not meant to apply to *de facto* rigid designators like "the number of primes between 10 and 20".)

Even with this restriction, problems arise for (CTR). That doctrine has a hard time dealing with cases like Kaplan's "Newman 1" and expressions rigidly (in both the de facto and de jure sense) denoting numbers and other abstract objects.

I myself think that the causal theory is wrong even for "Nixon" and

“Socrates.

But this is neither here nor there. What I wish to show now is that *even if* (CTR) is unrestrictedly true, there is still no opposition between it and (SC). There is no more opposition between those two views than there is between the view that a triangle is closed, planar, straight-edged figure and the view that a triangle is the area bounded by three coplanar lines such any two of them intersect, but not all three of them intersect.

Let us refer to advocates of CTR as “causal theorists”.

The causal theorist must say *which* causal relations constitute, or underwrite, reference. For not just any causal relation will do. If I push the door-bell, that causes the bell to ring. But the ringing of the bell doesn’t refer to my pushing of the button, at least not in quite the sense in which tokens of “Socrates” refer to Socrates. So the causal theorist must have a *specific* type of causal connection in mind.

There are many kinds of causal relations. And it is not immediately obvious which causal connection is the right one.

For the sake of argument, suppose that (CTR) is correct. Let R be the causal relation, whatsoever it should turn out to be, that constitutes or underwrites the relation of reference. In other words, R is the causal connection that causal theorists have in mind (or, better, that they are seeking).

Obviously R cannot be just *any* causal relation. A couple of extremely *general* constraints are obvious. For example, the causal theorist cannot say: R holds between E and O iff O gives rise to tokens of E in a way that allows the latter to refer to the form. That would be viciously circular.

But apart from such highly general constraints, it is not obvious what the right causal relation is. I am not saying that it doesn’t exist, or even that people haven’t figured out what it is. I am saying that *some* thought is needed

to know what R is.

How do we know if a given type of causal connection is the right one? What is our benchmark? Suppose some not very intelligent causal theorist says:

(*) Consider the causal relation that holds between pushing the button at t and the bell's ringing at t*. Let us refer to that relation as DB. R is the kind of causal relation that holds between Socrates (or, more exactly, states of affairs involving him) and tokens of "Socrates" – it is the relation that makes those tokens refer to that person.

Everyone, including most causal theorists, would *reject* (*). DB obviously does *not* do the job. If indeed tokens of "Socrates" stood in that relation to Socrates himself, that would *not* be enough for such tokens to refer to the man. Why not? By supposition, the ringing of the bell stands in DB to the pushing of the button. But the ringing of the bell surely doesn't *refer* to the pushing of the button. Perhaps the former *represents* or, at least, *indicates* the latter. Perhaps the former carries *information* about the latter. But the former doesn't *refer* to the latter. DB is, apparently, not a reference-constituting relation. So even if DB holds between Socrates and tokens of "Socrates", that is not the reason the latter refer to the former.

Why is DB deficient? Why is it so obvious that DB is the wrong relation? Given that we are so quick to say that DB is the wrong causal relation, we obviously have some pre-conception of what the right causal relation would have to do. We have some target in mind, some pre-theoretic conception of what reference is. And (*) is not hitting that target.

The target, I think, is this. If you want to make a statement *about* some

pushing of some button, you cannot use the resulting bell-ringing as a way of doing so. Suppose that the button is pushed at t, and the bell rings at t*. Let RING be that particular ringing. And let PUSH be the button-pushing in question. The sound-token:

(#) "[RING] [194] is a distal cause of the dog's waking up"

does *not* bear the proposition:

(##) PUSH is a distal cause of the dog's waking up.

(#) doesn't say *anything*. You have an unmeaning noise, followed by "is a distal cause of the dog's waking up". So (#) no more bears (##), or any other proposition, than does:

(###) "posidhfoasfhi is a distal cause of the dog's waking up"

What RING is missing is the ability to play a role in sentences. RING cannot be a part of a well-formed English sentence. That fact, by itself, bars us from saying that RING *refers* to anything, at least in English. So, at the very least, in order to qualify as a term that referred (in English) to PUSH, RING would have to be able to be a grammatically significant constituent of English sentences. It must make a semantically significant contribution to sentences.

But, though necessary, that is obviously not enough. The word "Morgan Freeman" make semantically significant contributions to English sentences. But "Morgan Freeman" doesn't refer to PUSH. The word "and" makes

semantically significant contributions to sentences. But “and” doesn’t refer to PUSH. “And” contributes the wrong thing.

If RING is to refer to PUSH, then it very much seems that, in addition to contributing *something* of significance to sentences, the thing it contributes must be PUSH itself.

Let us sum up. If an expression E is totally incapable of playing any significant role in sentences of a language, then E is not a referring term -- at least not a referring term of that language. So E must be able to contribute something to the semantics of sentences of a language if it is to be a referring term in that language. But for E to refer to some specific object O in L, it is obviously not enough that E make just *any* kind of significant contribution to L. E must contribute some specific thing. The natural choice here is: E must contribute O; L-sentences of the form “…E…” must mean:…*O*…

Once it is granted that a term must have *some* grammatically licensed role if it is refer, then it becomes hard to deny that (SC) gives at least more or less the right idea.

the relevance of this to CTR

None of this means that the CTR is wrong. That is not what I am trying to show. What I am trying to is that, when we are looking for the *right* casual relation – the causal relation that, according to CTR, enables E to refer to O - the constraint we have in mind is this: that causal relation must enable E to contribute O to sentences in which the former occurs. The reason RING doesn’t refer to PUSH (at least not in English) is, specifically, that the causal relation in question evidently doesn’t enable RING to semantically contribute PUSH. Utterances of the form “…[RING]…” don’t mean:…*PUSH*…

No one, not even a *causal* theorist, is going to say that RING refers to

PUSH. Even though RING has *some* causal relation to PUSH, it is not the *right* one. So what is the *right* causal relation? Supposing that CTR is right, what kind of causal theory *would* have to obtain between the RING and PUSH for the former to refer to the latter? The causal relation that *would* have to obtain is one that would allow RING to contribute PUSH to the senses of sentences in which RING occurs.

The causal theorist says that we'll know when we've identified the *right* casual relation – the reference-constituting relation – when we've hit on a causal relation R such that, if E stands in R to O, that is sufficient (and perhaps necessary) for E's having a certain *semantic* relation to O. What is that semantic relation? It is the one that holds between (tokens of) "Socrates" and Socrates. What is *that* relation? Well, "Socrates" is obviously capable of making a semantic contribution to English sentences. If it couldn't, then it surely would not be a referring term (at least not a referring term of English). And, what is also obvious, not just *any* semantic contribution will do; the semantically significant role just described must involve Socrates; it must play the role of introducing *Socrates* to the senses of sentences in which it occurs.

So the casual theorist does not say: "all this business about semantic contribution is irrelevant." On the contrary, he says (or should say):

"We'll know that we have the *right* causal relation – the one that holds between tokens of "Socrates" and Socrates – when we have a causal relation that is consistent with (SC). Tokens of "Socrates" have a certain causal connection to Socrates. And that is why they refer to Socrates. Now tokens of "Socrates" are able to do certain things. For example, such tokens are able to figure grammatically in English sentences. Obviously if "Socrates" couldn't figure grammatically in English sentences, then "Socrates" wouldn't refer to

anything in English. Further, such tokens contribute Socrates – not Plato or Charles Dickens – to the senses of sentences in which they occur. Obviously if "Socrates" contributed something other than Socrates, then such tokens wouldn't refer to Socrates. Tokens of Socrates have all of these wonderful properties in virtue of the fact that they have a certain causal relation to Socrates. Of course, not just any causal relation will do. We will know we have the right causal relation precisely when we've hit on a relation R such that, if tokens of E stand in R to O, then those tokens are able to do all the wonderful things that tokens of "Socrates" do in English – they can figure grammatically in English sentences and contribute O to the senses of those sentences.

The other side of the coin is this. *If* (per impossibile) the causal relation between RING and PUSH *were* such that RING semantically contributed PUSH, then causal theorists would no doubt say: R is the *right* causal relation (or, at least, *a* ***[195]*** right causal relation).

Once again, let DB be the causal relation that holds between PUSH and RING. Suppose (counter-conceptually) it turned out that, whenever one event E1 stood in DB with respect to some event E2, sentence-tokens of the form "…E1…" meant:…*E2*…In other words, sentences of the form "…E1…" would bear the proposition…*E2*…iff E1 stood in DB with respect to E2.

In that case, the causal theory would be strikingly vindicated; and causal theorists would surely say: "You see! Our theory is right! Reference *is* a causal relation. And that causal relation is DB."

I highly doubt that, under those circumstances, causal theorists would say:

So sentences of the form "…E1…" bear proposition…*E2*…iff E1 stands in

DB with respect to E2. That is fine and well. But it doesn't help *our* theory out. Our work is not done; we have yet to find the right causal relation. For reference is *not* the semantic-relation just described; it is something else. So given only that DB constitutes or realizes the semantic relation just described, nothing follows about whether DB is the right causal relation – the one that constitutes reference.

The basic idea is that (SC) *does* say in what reference consists, even by a causal theorist's lights. When he is considering the various different causal relations that hold between events, the causal theorist makes his selection on the basis of which ones are consistent with (SC). Suppose that causal relation ® can hold between an object O and tokens of some expression E *without* its being the case that those tokens semantically contribute O. In that case, everyone – including causal theorists – would say that ® was the wrong relation. So the causal theorist doesn't reject (SC). On the contrary, it is his touchstone, his compass.

We could, at least arguably, *invent* a language in which the ringing caused by a given doorbell pushing actually *referred* to that pushing. Let L be a language that is just like English except that L is characterized by the following semantic rule. In L, the ringing produced by a door-bell pushing actually *refers* to that doorbell pushing. Let *RING* be some particular door-bell ringing. In L, the spoken sentence-token "[*RING*][196] was a partial cause of my waking up" means that a certain door-bell pushing woke me up. [197]

Here we have a clear case where a causal relation is involved, in an essential way, in the reference relation.

But notice that the causal relation *by itself* does not suffice for the

reference relation. What suffices is that causal relation *plus* some semantic rule to the effect that iff x and y stand in that causal relation, *then* sentences of the form "…x…" are to mean:…*y*…After all, if we had the causal relation in the *absence* of that ad hoc semantic rule, then surely "…RING…" would quite definitely fail to mean:…PUSH…(In fact, in English, the former obviously *does* fail to mean the latter, specifically because English does not have that extra semantic rule.) So the causal relation in question is reference-constituting only in the rearguard sense that there is some semantic rule – and I might add, it is an extremely contrived one – stipulating that, when that causal relation obtains, a certain relation of semantic-contribution is to obtain.

So in the end, reference is to be understood in terms of the relation of semantic-contribution. And the causal theory is correct only in so far as some causal relation can be found that can perform the function of making this semantic contribution – the one described by (SC).

a note on CTR

This discussion, I believe, gives us some insight into what is i alleged – or can be coherently alleged – by any causal theory of reference. And this will, in turn, substantiate our view that SC gives the right analysis of reference.

We just said that the causal relation between RING and PUSH does not *by itself* enable RING to refer to PUSH. It was necessary to create an extra semantic rule. That extra semantic rule *exploited* the causal relation between RING and PUSH; for that rule was to the effect that RING should refer to that thing x, whatever x should turn out to be, that is causally related to RING in manner R (for some value of R). So it is really that semantic rule that enables RING to refer to PUSH; the causal relation is relevant only to the extent that some semantic rule makes it relevant.

I believe that something very close to this is true of the causal connection between tokens of “Socrates” and Socrates. *Part* of what makes such tokens refer to Socrates is that a certain causal connection holds. But that connection has a subordinate role. It is sub-ordinate to a certain semantic rule. And, strictly speaking, that *semantic* rule is what makes those tokens refer to Socrates. That rule may *exploit* a causal connection existing them and Socrates. But it is the rule, not the causal connection, that secures reference. In any case, this is what I will now argue.

Here is my argument. For the sake of argument, suppose that the causal theorist is right; suppose there is some causal relation R such that x (a given sound or inscription) refers to y iff y stands in R to x.

In that case, tokens of “Socrates” refer to Socrates because Socrates stands in relation R to those tokens.

Obviously there is an Albanian word for Socrates, a Finnish word for Socrates, an Urdu word for Socrates, and so on. These are[198] different words. According to the causal theorist, for any token T of any one of these words, what makes T refer to Socrates is that Socrates stands in R to T. So “Socrates” refers to Socrates because Socrates stands in R to tokens of Socrates. Let N be the Urdu for Socrates. According to the causal theorist, tokens of N refer to Socrates because Socrates stands in R to N. The same is true *mutatis mutandis* for the Albanian word Socrates, the Finish word, the Hungarian word, and so on. Given any token of any one of these words, that token refers to Socrates entirely because Socrates stands in R with respect tot hat token.

Let us focus on N, the Urdu word for Socrates. N is not “Socrates”; it is some other word.[199]

If you were to put N in an English sentence (I mean, if you were to *use* N

in an English sentence, not *mention* it), the result would be ungrammatical; you would not have a well-formed sentence of English. If you were to say *N was wise* or *N drank hemlock*, the result would not be a well-formed sentence. People might understand you – after all, communication doesn't always involve well-formed utterances. (If I child emits a certain kind of cry, I know that it wants food or wants its diaper to be changed. But its cry is not English or Urdu or any other language.) When you said *N was wise*, they might say: "You mean *Socrates* was wise." Their understanding you would involve their doing a bit of translation. This means that your utterance was not a well-formed utterance of *English*; it contained a renegade constituent, one that had to be replaced with an English word to yield proper English.

The important point is that English sentences of the form *…N…* are not well-formed. N is not a word of English.

Given any referring term, it will be a referring term *of* this or that language. It will not be a referring term *simpliciter*. N refers to Socrates in Urdu, not in English. "Socrates" refers to Socrates in English, not in Urdu.

If the causal theory is right, then what makes tokens of N refer to Socrates is wholly that Socrates stands in relation R to those tokens. After all, the causal theorist says that x refers to y iff y stands in relation R to x. N stands in relation R to Socrates. Therefore N refers to Socrates.

But *in English* N obviously does *not* refer to Socrates. Quite possibly, in any language other than Urdu, N does not refer to Socrates.

But the causal theorist says that what makes tokens of "Socrates" refer to Socrates is simply that the latter stands in R to the former; the reference-relation in question is *constituted by* that causal relation. That causal relation is quite enough, according to the causal theorist, for "Socrates" to refer to Socrates in English. But the problem is that tokens of N stand in that relation to Socrates no less than tokens of "Socrates". And yet tokens of N clearly *do*

not refer to Socrates in English. Clearly, then for x to refer to y in some language L, it is *not* enough that y stand in R to x. That may be a part of the story; but it cannot be the whole of it, as we've just seen.

The causal theorist might resist this; he might say this:

The causal relation that an *Urdu* word must bear to Socrates if that word is to refer to Socrates is different from the causal relation that an *English* word must bear to Socrates if it is to refer to Socrates. And both of those relations are different from the causal relation that a *Tagalog* word must be to Socrates if that word is to refer to Socrates. N bears the casual relation to Socrates that is necessary for N to refer to Socrates *in Urdu*, but not in English. "Socrates" (or, more accurately, tokens thereof) bears the causal relation to Socrates than is necessary for an *English* word to refer to Socrates.

But it is clear what the problem is here. This position disjunctives the concept of reference. There is referenceinUrdu, referenceinEnglish, referenceinFrench, and so on.[200] But there is no longer a *single* concept of reference. And this is plainly unacceptable. For N refers to Socrates in the very sense in which "Socrates" refers to him, and in which the Chinese word for Socrates refers to him.

If you *identify* reference with a causal relation, and *then* say that one causal relation has to hold between Urdu words and Socrates if they are to refer to him, and a different relation must hold between *English* words and Socrates if they are to refer to him, then you are saying that the Urdu word for Socrates does not refer to that man in quite the sense in which the English word refers to him.

But this is false. The Urdu word refers to Socrates in exactly the sense in

which the English word does. So it is not open to the causal theorist to say that *different* causal relations constitute reference, depending on the language. He must say that some *one* relation R does the job. So, if the causal theory is right, then N and "Socrates" both stand in R with respect to Socrates; and it is exactly this which makes both N and "Socrates" refer to that philosopher. But, as we've seen, this position is untenable.

At this point, the causal theorist may weaken his position and say:

Maybe you're right. To avoid falsely disjunctivizing the concept of reference, one must say that reference is not *identical* with some causal relation. Very well. I never said that reference *is* some causal relation; I said that a certain causal relation is essentially *involved* in reference. So reference *per se* is not some causal relation; it is something else. That something else *involves* some kind of causal relation. But reference itself isn't quite the same thing as this causal relation.

But then we are left with the question: what is reference? Let us start out by conceding as much as possible to the causal theorist. Let us suppose that there *is* some causal relation R that is essentially involved in all cases of reference. With regard to "Socrates", and N, and all the other words that refer Socrates, let us suppose that Socrates stands in R to tokens of those words, and that this is at least *part* of the reason that those tokens refer to Socrates.

We've seen that for x to refer to Socrates, it isn't *sufficient* that Socrates stand in R to x. Something must be added. What is this something? What makes "Socrates" refer to Socrates in English, besides the fact that Socrates stands in R to tokens of "Socrates"? Whatever that extra something is, it must suffice to make "Socrates" be capable of *grammatical* use in English sentences. If a word w cannot be *grammatically* used in English, then it

surely doesn't refer to anything in English (though it may refer to something in Arabic or Basque). As I see it, there is only one thing that can make a physical object figure in a grammatically acceptable way in a sentence of some language: some *semantic rule* belonging to that language must give that object a meaning within that language.

This is a truism, but a relevant one. If some sound or inscription w is to have any kind of meaning in English, some semantic rule of English must exist that gives it meaning. Obviously if (say) a sound is to refer to Socrates in English, then there must be some semantic rule of English to that effect. And if there is no such rule, then it won't refer to Socrates in English.

Suppose somebody said this:

In English, "Socrates" refers to Socrates. But English does not comprise the following semantic rule: *"Socrates" refers to Socrates.*

That would be nonsense. Wherever there is reference, there is a semantic rule. This is a truism.

We've seen that no causal relation *by itself* constitutes the reference-relation.

Consider *any* one of the causal relations that Socrates bears to tokens of "Socrates": no one of those causal relations suffices to make those tokens refer to Socrates in Urdu or Chinese or Finnish. By parity of reasoning, they are not what suffices to make those tokens refer to Socrates in English.

We've also seen that, if a word is to refer to anything in a language, then there must be some semantic rule of that language assigning that referent to that word.

What room is there left for the casual theory? There is good reason to believe that a casual connection between tokens of "Socrates" and Socrates

plays a *part*[201] in the fact that the former refer to the latter. Kripke made an excellent case for this. But we've also seen that no causal relation *suffices*; and we've also seen that, if English has the right kind of semantic rule, that *does* suffice.

finding room for the causal theory

Here is where the casual theory becomes relevant. Consider the English semantic rule that is involved in the fact that "Socrates" refers to Socrates. That rule may well *exploit* some causal connection obtaining between certain current sound- and inscription-tokenings, on the one hand, and states of affairs involving Socrates, on the other. That rule may be at least vaguely like:

(CRS) Let s be an arbitrary token of the physical type "Socrates". If there is some x such that x uniquely stands in causal relation R [for the appropriate value of R] with respect to s, then s refers to Socrates. If there is no such thing – if nothing stands in R with respect to s, or multiple things do – then s refers to nothing.

So given (CRS), and given that Socrates is the thing standing in relation R with respect to my tokening of "Socrates", it follows that Socrates is what that tokening refers to. If Plato were the thing standing in that relation to that tokening, that that tokening would refer to Plato.

The point is that it is not the causal relation *per se* that gets tokens of "Socrates" to refer to Socrates; it is some semantic rule which uses that causal relation.

Now we can begin to close this argument.

Given any semantic rule that makes Socrates be the referent, in English, of some sound or inscription, any such rule must make tokens of the physical type "Socrates" be capable of figuring *grammatically* in English sentences. If x refers to y in L, then (at the very least) x must be capable of having a grammatical role in sentences of L.

So if you say that x refers to y in L, you are saying, among other things, that x can grammatically figure in sentences of L.

When we say that x can figure *grammatically* in sentences of L, we are saying, surely, that x can make some kind of semantic contribution to the meanings of sentences of L. So when you say that "Socrates" refers to Socrates in English, you are saying that "Socrates" makes some kind of contribution to the meanings of sentences of English.

What is that contribution? Socrates is that contribution. Sentences of the form "…Socrates…" mean:…*Socrates*…

an application

Given what we've said, it is pretty clear that a causal relation cannot, by itself, *constitute* reference. Causal connections are involved in reference only when semantic rules exploit them.

This insight enables us to circumvent some problems associated with CTR. We can obviously refer to things to which we have no causal connection – future individuals, abstract objects, and possibly certain contemporary spatio-temporal objects.

But there is only one concept of reference. "2" refers to the number two in exactly the sense in which "Nixon" refers to Nixon. There is no causal connection between the first pair. So CTR runs the risk of disjunctivizing the concept of reference. (Here we are dealing with a *different* disjunctivization

of reference from the one discussed before. So CTR is guilty of doubly disjunctivizing an obviously unified concept.)

Advocates of CTR have tried to work around this, saying things like: "2" is causally connected to *instances* of the number two. But such maneuvers are obviously implausible, and don't even appear to work when we're dealing with future individuals.

These problems vanish when CTR is put in its proper context.

Reference is always secured in the same way. We *describe* the thing to which an expression is meant to apply. But we give the description wide-scope in the semantic rule we are inventing. That way the sole semantic content of the expression is the individual: the expression doesn't become a Russellian description-quantifier.

Suppose I want to produce a term that refers to the first individual born in the 3rd millennium. I say:

(3M) *If* there is some object O such that O is the first individual borne in the 3rd Millennium, then "Simon" refers to O.

Because the descriptive content in that rule is given wide-scope with respect to the "refers to" operator, the statement: "There are worlds where Simon is born in the 2nd millennium" are true. And this is what we want: for any individual *could* have had a slightly different birth-date.

As we've seen, in so far as CTR has any chance of being right, it is because certain semantic rules *exploit* causal connections. If a causal connection between tokens of "Socrates" and Socrates is involved in the fact that the former refer to the latter, it is because some rule like the following

exists:

(S) If there is some individual O such that tokens of “Socrates” stand in R to O, then those tokens refer to O.

(S) and (3M) are completely parallel. In both cases, reference is secured through descriptive information. But that information is given wide-scope with respect to the “refers to” operator. Where (S) is concerned, the descriptive content has to do with causal relations; where (3M) is concerned, this is not so. But that doesn’t have any special ramifications. Descriptions differ; some involve the concept of causality; some don’t. Some involve the concept of wearing a feathered-hat; some don’t. I can describe an entity in terms of its causal liasons or in terms of what kinds of clothes it is wearing. There are only two important facts: (i) in semantic rules, at least those concerned with the assignment of a referent to a term, descriptive information is used; and (ii) that information is given wide-scope with respect to the “refers to” operator. CTR merely registers the banal fact that, in some cases, the descriptive information in question has to do with a thing’s causal liasons.

Once this is realized, we don’t have to invent tortured hypotheses as to why there really is a causal connection between the number two and tokens of the “2”. There is no such connection. And even if there were, it wouldn’t be reference-constituting. What would be reference-constituting would be a semantic rule that fixed the referent of “2” in terms of descriptive information. That rule might exploit a description that alluded to the causal liasons of certain things – just as it might exploit a description that alluded to the numerical properties of a thing. But beyond that, the causal connection between the number two and tokens of “2” would not be reference-

constituting.

conclusion

We have substantiated our thesis about reference. And we found our theory not to be in conflict with the casual theory. Indeed we found it to constitute a needed addition to the causal theory. For “Socrates” to refer to Socrates is for “Socrates” to make a certain kind of semantic contribution to the meanings of sentences in which it occurs. And that contribution is given by the proposition: “…Socrates…” means:….*Socrates*…, for any predicate “…x…”

Chapter 7. are quantifiers referring terms?

In the argument we gave earlier, we assumed that quantifiers are *not* referring terms. This is a natural assumption to make. For reasons Frege and Russell gave, it is beyond dispute that “some man” does not refer to some man, that “nothing” doesn’t refer to some mysterious individual (contrary to what Plato thought). These terms function not by picking out curious individuals, but by “contributing their quantificational structure”[202]. So, it is natural to say, these terms *do not* refer.

But given only that quantifiers don’t refer to individuals, it doesn’t follow that they don’t refer at all. Some – Frege[203], Montague[204] – hold that they refer to functions.

In my argument I say repeatedly that quantifiers are not referring terms.

Definite descriptions, when functioning attributively, are quantifiers. Thus, when functioning attributively, they *are* referring terms. This seems to destroy my argument. For I suppose that quantifiers are not referring terms.

Here is what I would like to show. It may well be that quantifiers *are* referring terms. (In fact, I myself think they are.) But that doesn't matter as far as our argument is concerned. Suppose that "the inventor of bifocals" and "the first post-master general" are quantifiers, and suppose that quantifiers *are* referring terms. Given those two suppositions, those expressions are not *co-referring terms*, even though they are referring terms. In general, given the two suppositions just stated, *any* two definite descriptions that encode *different* descriptive content will *not* be *co*-referring terms.

And that destroys the Frege-Church argument. That argument requires that, in replacing (for example) "the inventor of bifocals" with "the first post-master general", one is replacing a referring term with a *co*-referring term. In general, that argument assumes that, in replacing "the phi" with "the psi" (where phi and psi are different properties, and where some one individual x uniquely has both phi and psi), one is replacing a referring term with a *co*-referring term. But, I will now argue, if "the inventor of bifocals" and "the first post-master general" are quantifiers that refer, then they do not *co*-refer. In general, if "the phi" and "the psi" are quantifiers that refer, they do not *co*-refer, (except in the irrelevant case where phi is the very same descriptive information as psi). And that nullifies the Frege-Church argument.

Let me briefly say why it is that *if* "the inventor of bifocals" and "the first post-master general" are terms that refer to functions or properties, they refer to *different* functions or properties.

Let w be this world, and let w* be some other world where English is spoken. Suppose that, in w*, Benjamin was the inventor of bifocals, but not the first post-master general. In that case, in w*, the sentence (or, to be

precise, sentence-token):

(*) “Benjamin Franklin was the inventor of bifocals”

is true, but th sentence

(**) “Benjamin Franklin was the first post-master general”

is false.

By assumption, those sentences have the same semantics in w* that they have here. So supposing that definite descriptions denote functions or properties, the definite descriptions in (*) and (**) denote the very functions (or properties) that they denote here. those sentences are exactly the same *apart* from the one fact that, in the place where the one has “the inventor of bifocals”, the other has “the first pos-master general”. In w*, (*) and (**) different truth-values. The only explanation is that the function denoted by “the inventor of bifocals” is *different* from the function denoted by “the first post-master general”. They do not denote the same function; they do not co-refer.

Let me now say a bit more on this matter.

quantifiers as referring terms

If quantifiers are referring terms, what do they refer to? The following gives the general idea. Consider the sentence:

(*) for some x, x is bald.

One the face of it, the quantifier – “for some x” – doesn’t refer to anything. But a case can be made that it refers to a second-order property. For (*) is equivalent to:

(**) <x is bald> is instantiated.

<x is bald> is a sequence whose sole member is the open-proposition *x is bald.*

Thus, the “for some x” in (*) turns out to refer to the property of being uniquely instantiated.

Another more traditional way of dealing with “for some x” is as follows.

First of all, by a “concept”, I mean a certain kind of abstract entity – the kind of thing that is expressed by predicates (e.g. “bald”, “fat”). I don’t mean some psychological entity.

A concept (e.g. the concept *x is bald*) can be seen as a function from objects to truth-values. (So it assigns truth to Patrick Stewart and falsity to Alec Baldwin.)

So the quantifier “for some x” can be seen as a function from concepts to truth-values. So it assigns truth to the concept *x is bald* iff it is instantiated, and falsity otherwise.

definite descriptions as quantifiers that refer

Supposing that definite descriptions are quantifiers that refer, what do they refer to? There are a couple of possible answers. I will start with some traditional answers and then give a non-traditional one.

Like “for some x”, and “for all x”, “the phi” can be seen as a function from

concepts to truth-values. “The phi” denotes a function that assigns *truth* to a class C iff nothing falling outside of C has phi and exactly one thing falling in C has phi.

Consider “the inventor of bifocals snored”. Supposing that definite descriptions are referring-quantifiers, this sentence is true iff exactly one thing that is a bifocal-inventor falls in that class and nothing that is a bifocal-inventor falls *outside* that class. So “the inventor of bifocals” can be seen as denoting a function from classes to truth-values. It assigns truth to a class C iff exactly one thing that is a bifocal inventor falls in C and nothing that is a bifocal inventor falls outside of C.

For the sake of argument, let us suppose that this is the right analysis of “the phi”. In that case: “the inventor of bifocals” denotes a function – call it F -- that assigns truth to a class C iff exactly one thing in C invented bifocals and nothing outside of C invented bifocals.

And “the first post-master general” denotes a function – call it F* -- that assigns truth to a class C iff exactly one thing in C was a first post-master general and nothing outside of C was a first post-master general.

the problem for our analysis

Here is the apparent problem for our analysis. It seems that F and F* are the very same function. K and K* are the same class iff they have exactly the same members. The class of bifocal inventors is therefore identical with the class of first post-masters general. (The properties in question are different, but not the classes they generate.) It follows that F and F* are the same function. Thus, if definite descriptions are referring-quantifiers – and they may well be -- then “the first post-master general” and “the inventor of bifocals” *do* co-refer if they refer to quantifiers. But my argument requires

that they *not* co-refer if they are quantifiers. So my argument fails, it seems.

the reason that argument doesn't affect our position

In my view, this argument involves a serious non-sequitur. Let w be this world and let w* be some other world where the *semantics* of English is just what it is here in w. But suppose that in w* Thomas Jefferson was the first post-master general. Obviously "the inventor of bifocals" and "the first post-master general" have the very same semantics in w* that they have in w. But if what was said in the last paragraph is correct, it is immediately follows that, in w*, "the inventor of bifocals" and "the first post-master general" denote totally different functions, while in w they denote the same function. But this is not possible: for, in w*, the *semantics* of those expressions is exactly what it is in w.

As we just saw, if quantifiers are referring terms, then "the inventor of bifocals" assigns truth to a class iff that class contains exactly one member of the class of bifocal inventors and no member of the class of bifocal inventors falls outside of C. And "the first post-master general" assigns truth to a class iff that class contains exactly one member of first post-masters general and no member of the class of first post-masters general falls outside of C.

Let K1 be the class of bifocal-inventors in this world. Let K2 be the class of first post-masters general in this world

K1 and K2 coincide.

Let K*1 be the class of bifocal-inventors in w1. And let K*2 be the class of first post-masters general in w*.

K*1 and K*2 do not coincide.

By stipulation, the semantics of definite descriptions in w* is exactly what it is here. (This is not a tendentious stipulation; there is nothing wrong with

supposing that there are possible worlds where people speak English.[205]) But if we analyze definite descriptions in the way proposed, then their semantics *does* vary: "the first post-master general" denotes one function in w – a function involving K1, which is identical with the singleton containing Benjamin Franklin. In w*, "the first post-master general" denotes a function involving K*1, which is the singleton containing Thomas Jefferson.

So "the first post-master general" denotes one function in w, and a different function in w*. Thus it has one semantics in w and a different one w* -- even though the semantics of English in w* is just what it is in w.

This is not acceptable.

The non-sequitur in the analysis in question is this: that analysis assumes that definite descriptions are functions involving classes in *extension*, as opposed to classes in *intension.* If "the inventor of bifocals" denotes a function involving classes, the classes in question cannot be ones that vary depending on who happens to have invented bifocals (or on whether they have a unique inventor or no inventor). Uncontroversially, a person can know the semantics of "the inventor of bifocals" (the expression *type*) perfectly well without knowing that it is a function involving a class that contains *Benjamin Franklin.* There are many people who understood that term and had no idea that Benjamin Franklin invented bifocals. (I myself used to be one of them.)

So if "the inventor of bifocals" is a term that refers to functions involving classes, they must be classes in *intension*. (They must be *properties*.) The result is this. If "the inventor of bifocals" is a term that denotes a function, then it denotes:

(IB) The function that assigns truth to a class C iff there is exactly one individual in the intersection of C and the class (intensionally defined) of

bifocal inventors and nothing outside that intersection is a bifocal inventor.

And if “the first post-master general” denotes a function involving concepts, then it denotes:

(FP) The function that assigns truth to a class C iff there is exactly one individual in the intersection of C and the class (intensionally defined) of first post-masters general and nothing outside that intersection is a post-master general.

(IB) and (FP) are not the same function. Obviously there are possible worlds where one person is a snorer and uniquely invented bifocals, and where some other person is a non-snorer and was a unique first post-master general. In such a world, (IB) assigns truth to the class of bifocal-inventors who snore, whereas (FP) assigns falsity to that same class. So plainly (IB) and (FP) are not the same function.

This is consistent with our intuitions. There are worlds where people speak English exactly as it is spoken here, and where the sentence “the first post-master general snored” is true and “the inventor of bifocals snored” is false. In such a world, those definite descriptions have the same semantics that they have here. But in that world, those definite descriptions obviously assign *different* truth-values to the same class. So obviously those definite descriptions denote *different* functions (supposing, indeed, that they denote functions at all).

There is a general lesson here. If quantifiers denote functions from classes to truth-values, the classes in question must be understood intensionally. Otherwise the meaning of the quantifiers will be contingent on how many men, penguins, and cars there are.

Summary

Definite descriptions can be seen as expressions that pick out functions from classes to truth-values. If that is the right view, then "the inventor of bifocals" and "the first post-master general" do *not* co-refer. Let us generalize this point. Suppose that there is exactly one thing x that is a phi and one thing y that is a phi and anything that is a phi is a y. (to avoid complications, suppose that phi and psi are different properties.) In that case, if definite descriptions are terms that refer to functions from classes to truth-values, "the phi" and "the psi" do *not* corefer.

Another analysis of definite descriptions

The conclusion I want to establish is that, *if* definite descriptions are quantifiers, then "the phi" and the "psi" do *not* corefer. Obviously what we've said gets us closer to that result. But we are not there yet. For, supposing that definite descriptions are quantifiers that refer, they don't *have* to be analyzed as expressions that denote functions (though that is perhaps the best way to analyze them); they can also be analyzed as expressions that denote properties of a certain kind.

Definite descriptions can be seen as things that pick out second order properties. Thus "the inventor of bifocals" can function as something that picks out a second order property, and so can "the first post-master general."

I will show that *when* those expressions are functioning as things that pick out second-order properties, they pick out *different* second-order properties. (So, once again, we will have the result that, if definite descriptions are

quantifiers that refer, then “the first post-master general” and “the inventor of bifocals” pick out *different* things and are thus *not* co-referring terms.)

When “the inventor of bifocals” is treated as a quantifier, “…the inventor of bifocals …” means this:

(i) Something x was uniquely an inventor of bifocals and…x…

So “the inventor of bifocals smoked” means:

(ii) Something x was uniquely an inventor of bifocals and x smoked.

Supposing that the quantifier in (**) is a referring term, as the objector maintains, what second-order property does it refer to? To answer this, we must reconfigure (ii). (For the form of (ii) doesn’t make it perspicuous what the relevant second-order property is.)

(iii) < *x smoked*> is instantiated and is such that anything instantiating it uniquely invented bifocals.

(iii) is, of course, equivalent to (ii). And, in (iii), the *quantificational structure* of “the inventor of bifocals” is seen to consist in its picking out a certain property, namely: *is instantiated and is such that anything instantiating it uniquely invented bifocals.* In other words, when “the inventor of bifocals” is functioning as a quantifier, its function *is* to pick out a certain second-order property.

So the objector does seem to be right. Definite descriptions, read attributively, are referring terms: they pick out second-order properties.

But when “the inventor of bifocals” and “the first post-master general” are treated as terms that refer to second-order properties – when they are treated as quantifiers – they pick out *different* second-order properties.

Read attributively,

(iv) The first post-master general smoked

means

(v) Something x was uniquely a first post-master general and x smoked.

To make it clear what the relevant second-order property is, we must reconfigure (v). The needed reconfiguration is not hard to make:

(vi) <*x smoked*> is instantiated and is such that anything instantiating it uniquely was a first post-master general.

So, when it is functioning as a quantifier, “the first post-master general” picks out the property: *is instantiated and is such that anything instantiating it uniquely was a first post-master general.*

But the property

is instantiated and is such that anything instantiating it uniquely was a first post-master general

is very different from the property

is instantiated and is such that anything instantiating it uniquely was an inventor of bifocals.

The first is the property picked out by “the first post-master general” (when that expression is read attributively). The second is the property picked out by “the inventor of bifocals” (when that expression is read attributively).

Let us now close the argument. The objector says that quantifiers are referring terms. This may well be true. Thus “the inventor of bifocals” and “the first post-master general”, when functioning as quantifiers, are terms that refer to second-order properties. But *when* they function in that way, they pick out altogether *different* properties. Thus, when functioning as quantifiers that pick out properties, “the inventor of bifocals” and “the first post-master general” pick out altogether different things: they are not *co*-referring terms.

Summary

Suppose that exactly one thing has phi, and exactly one thing has psi, and that any thing having phi has psi. Also, suppose that phi and psi are different properties (so that “the phi” and “the psi” are not mere terminological variants of each other).

If “the phi” and “the psi” definite descriptions are quantifiers that refer to properties, they refer to different properties.

If they are quantifiers that refer to functions, then they refer to different functions.

Given this, we must conclude that, if they are quantifiers that refer, then “the phi” and “the psi” refer to *different* things.

Chapter 8. the anti-slingshot rehabilitated

Let us begin with an overview of what we've just argued for.

If definite descriptions are terms that refer to individuals, then "…the inventor of bifocals…" means the very same thing as "…the first post-master general…" If "the first post-master general" and "the inventor of bifocals" are treated as quantifiers then they either do not refer at all *or* they refer to functions of some kind of properties of some kind. We've seen that, if they refer to functions, they refer to *different* functions. And we've seen that if they refer to properties, they refer to *different* properties. Given this, we must conclude that, if definite descriptions are quantifiers that refer, then "the inventor of bifocals" and "the first postmaster general" refer to *different* things. In general, , if definite descriptions are quantifiers that refer, then "the phi" and "the psi" refer to different things (except in the special case where phi and psi are the same property).

These facts demolish the arguments given by Frege and Church. Frege says (these are not his exact words):

You can change the sense of a sentence by replacing referring terms with co-referring terms. "The inventor of bifocals smoked" has a different sense from "the first post-master general smoked". But the second is what results when a referring term in the first is replaced with a co-referring term.

If the definite descriptions in those sentences are treated as *quantifiers*, then they are *not* co-referring terms. Either they are terms that refer to entirely different things or (as Russell thought) they are terms that don't refer at all.

If the definite descriptions in those sentences are *not* treated as quantifiers, then each must be treated as term that contributes an individual, and nothing

besides, to the proposition borne by the sentence in question. If “the inventor of bifocals” and “the first post-master general” are so treated, then they both pick out the very same individual. In that case,

(vii) “the first post-master general smoked”

and

(viii) “the inventor of bifocals smoked”

bear the exact same proposition, namely:

(ix) *Benjamin Franklin smoked.*

So it is *not* possible to change propositions by replacing referring terms with co-referring terms. Consider (vii) and (viii). The definite descriptions must be read either as quantifiers or not as quantifiers. If they are read as quantifiers, then (if they refer to anything at all), they pick out altogether *different* things. If they are *not* read as quantifiers, then they pick out the same thing, and the sentences in question bear the same proposition.

Either (vii) and (viii) bear the same proposition *or* neither results from the other by replacing referring terms with co-referring terms.

In general, it is not possible to induce a propositional-shift merely by replacing referring terms with co-referring terms.

It is clear how this point destroys Church’s argument. That argument assumption that propositional-shifts can be induced by replacing referring terms with co-referring terms.

The anti-slingshot rehabilitated

Now we can put forth an *anti*-slingshot that involves only assumptions that we have shown to be correct.

Let us consider Church's version of the slingshot (what we will say generalizes to the other versions). The core of that argument is given by the following sequence of sentences.

(i) Sir Walter Scott is Sir Walter Scott.
(ii) Sir Walter Scott is the author of Waverly.
(iii) Sir Walter Scott is the man who wrote twenty-nine Waverly novels altogether.
(iv) The number n, such that Sir Walter Scott wrote n Waverly novels, is twenty-nine.
(v) The number of counties in Utah is twenty-nine.

Church thinks that, given (CR), and given the principle that (nearly) synonymous expressions co-refer, it follows that (i) and (v) co-refer.

Church is wrong.

The definite descriptions can be read either referentially or attributively. When I talk about "referential" readings of definite descriptions, I mean readings on which they refer to *individuals.*

If a referring term is read referentially, then its sole semantic contribution is its referent. We have justified this claim.

If a definite descriptions is read attributively, then it is a quantifier. We have justified this claim. Contrary to what Frege and Mill and others thought, when a definite descriptions contributes its *sense,* then it must be functioning

as a quantifier, not as a referring term.

If a definite descriptions is read attributively, i.e. as a quantifier, then it is not a referring term or it is a term that refers to a function of some kind or a second level property.

Obviously the definite descriptions Church's argument must be read consistently. Otherwise, that argument will involve an equivocation-fallacy. We cannot give (say) "the number of counties in Utah" a referential reading in one line and an attributive reading in the next.

So let us consider all the possible readings of Church's argument. Let us start with the one wherein the definite descriptions are read referentially, i.e. as terms that refer to *individuals* and where their sole semantic contributions are their referents. In that case, Church's argument becomes:

(i_R) Sir Walter Scott is Sir Walter Scott.
(ii_R) Sir Walter Scott is Sir Walter Scott.
(iii_R) Sir Walter Scott is Sir Walter Scott.
(iv_R) Twenty-nine is twenty-nine.
(v_R) Twenty-nine is twenty-nine.

Read this way, Church's argument fails. The third and fourth sentences are not synonymous or "very nearly so".

So if Church's argument is to work, we must read the definite descriptions as quantifiers. There are two sub-options here. We can treat quantifiers as things that don't refer at all, or we can treat them as things that refer to functions or properties of some kind.

Every single step in Church's argument obviously relies on the assumption that definite descriptions are terms that refer to individuals. For example, it essentially involves the assumption that "Sir Walter Scott" and "the author of

Waverly" and "the unique author n such that n wrote twenty-nine Waverly novels" all co-refer, and refer to an individual. So if Church's argument is to work, we cannot treat the definite descriptions in it as terms that don't refer at all.

But, as we saw a moment ago, we cannot treat them as terms that refer to individuals.

So we must treat them as terms that denote functions (from concepts or classes to truth-values) or, perhaps, as 2nd level properties.

But in that case, the argument immediately fails. That argument assumes that "Sir Walter Scott" and "the author of Waverly" refer to the same thing. But if we treat definite descriptions as terms that refer to functions or second level properties, then those expression do not corefer; for "Sir Walter Scott" does not refer to a second level property or a function.

Also, more importantly, if we treat definite descriptions in that way, then no two of the definite descriptions in that argument co-refer. (For the remainder of this paragraph, assume that definite descriptions denote functions or second level properties.) Consider a world w* where English is spoken, where Scott wrote *Waverly*, but didn't write any other Waverly-novels. In w*, "some human was the author of *Waverly"* is true, whereas "some human was the author of twenty-nine Waverly-novels" is false. So, in w*, "the author of Waverly" assigns truth to the class of humans, whereas "the author of twenty-nine Waverly-novels" assigns falsity to that same class. So the function denoted by the one definite descriptions is apparently distinct from the one denoted by the other.

The same is true *mutatis mutandis* of any two of the definite descriptions in Church's argument. No two of them co-refer, if they are taken as quantifiers.

But Church's argument essentially requires that some of the co-refer –

that, for example, “the author of Waverly” and “the author of twenty-nine Waverly novels”.

So Church’s argument fails. And it fails on *any* reading of definite descriptions.

The reasons that Church’s argument fails are not peculiar to Church’s specific version of the slingshot. All versions of the slingshot require that replacing a descriptions with a co-referring definite descriptions can change result in a change in a sentence that encodes a new proposition. All versions of the slingshot involve the assumption (here I am stating it in a specialized form) that

(a) “the inventor of bifocals snored”

and

(b) “the first post-master general snored”

encode different propositions, even though “the inventor of bifocals” and “the first post-master general” co-refer.

In general, all versions of the slingshot involve the assumption that, for any predicate “…x…”, and for at least some values of “phi” and “psi”,

(c) “…the phi…”

and

(d) “…the psi…”

encode different propositions, even if “the phi” and “the psi” are co-referring terms.

But this assumption is false. If “the phi” and “the psi” are co-referring terms that refer to individuals, then (c) and (d) will encode precisely the same proposition. If “the phi” and “the psi” are quantifiers, then one of two things is the case. The first case is this: “The phi” and “the psi” don’t refer to anything at all (this is what Russell thought). In that case, of course, “the phi” and “the psi” do not co-refer.

The second case is this: “the phi” and “the psi” *do* refer to functions or second-level properties. But they refer to *different* functions (except in the irrelevant case where “the phi” and “the psi” are actually identical or synonymous – where, for example, “phi” is the description *first-poster general* and “psi” is the description *first post-master general* or *person x such that nobody before x held the office of postmaster general and such that x occupied that post*). Once again, the assumption in question turns out to be false.

A Salmon-Kripke reply

I must consider an important response to my argument (due to Nathan Salmon):

Here is a response to your argument: never mind English definite descriptions. Instead we introduce an artificial device “ze” that works according to the following rule: *ze phi*[206] refers to the unique individual that satisfies (has) phi, if there is such an individual, and fails to refer otherwise. In that case “Scott” and “ze author of Waverly” are co-referential. Now replace “the” through Church’s (or Frege’s) argument with “ze”. The

resulting argument is immune to what you've said.

This is, I honestly believe, about as good a response as could be made on behalf of the slingshot. Let us see how well it fairs.

(i) Sir Walter Scott is Sir Walter Scott.
(ii) Sir Walter Scott is ze author of Waverly.
(iii) Sir Walter Scott is ze man who wrote twenty-nine Waverly novels altogether.
(iv) Ze number n, such that Sir Walter Scott wrote n Waverly novels, is twenty-nine.
(v) Ze number of counties in Utah is twenty-nine.

Of course, "ze author of Waverly" contributes the very same thing as "Scott". They both contribute Scott. So the proposition encoded in (ii) is identical with the proposition encoded in (i).

"Ze man who wrote twenty-nine Waverly novels altogether" contributes Scott. So (iii) encodes the same proposition (ii) and as (i).

"Ze number n, such that Sir Walter Scott wrote n Waverly novels" refers to/contributes the number twenty-nine. So (iv) encodes the proposition: twenty-nine is twenty-nine.

"Ze number of counties in Utah" refers to/contributes the number twenty-nine. So (v) encodes the proposition: twenty-nine is twenty-nine.

So when we rewrite Church's argument, replacing "the" with "ze", and consider the propositions encoded in the resulting sentences, the result is this:

(i) Scott is Scott.

(ii) Scott is Scott.
(iii) Scott is Scott.
(iv) Twenty-nine is twenty-nine.
(v) Twenty-nine is twenty-nine.

Church's argument depends on the near-synonymy, or outright synonymy, of (iii) and (iv). Obviously that synonymy isn't there in the '"ze"-version. So replacing "the" with "ze" does not, at least not on the face of it, rehabilitate Church's argument.

For the same reason, it doesn't look as though replacing "the" with "ze" will much help Frege's argument. Frege tries to show that, given any sentence S, and any proposition P, a sentence S* that encodes P can be produced merely by replacing referring terms in S with co-referring terms (and, perhaps, reiterating this process). So given "Plato smoked", we can produce a sentence that encodes *arithmetic is incomplete* by replacing referring terms with co-referring terms: "the unique x such that x wrote The Republic and arithmetic is incomplete smoked". We've seen the problem with this gambit. If the definite description is interpreted referentially, then the just mentioned sentence encodes only the proposition *Plato smoked* and does *not* encode the proposition *arithmetic is incomplete.* If the definite description is interpreted attributively, then the word "Plato", in "Plato smoked" is being replaced with a *quantifier*, not a referring term. And, once again, Frege doesn't get the result he wants.

Suppose, however, that we replace "Plato" with "ze unique x such that x wrote The Republic and arithmetic is incomplete". The result is

(*) "ze unique x such that x wrote The Republic and arithmetic is incomplete smoked".

Bu, given the way “ze” is defined”, (*) bears the proposition

(**) *Plato smoked*.

So the proposition *arithmetic is incomplete* drops out. And, once again, Frege’s slingshot is blocked.

Summary

Church and Frege both argue that sentences refer to truth-values. Their arguments both assume (at least by implication):

(SS) It is possible to change the proposition borne by a sentence by replacing referring terms in that sentence with co-referring terms.

SS is false.

Suppose that phi and psi are different properties. Further, suppose exactly one thing has phi and that exactly one thing has psi, and that anything having phi also has psi.

Under that circumstance, there is a natural reading of “…the phi…” and “…the psi…” whereby they bear different propositions. But on that reading “the phi” and “the psi” are not functioning as terms that refer to individuals; they are functioning as terms that either do no refer at all or refer to *different* things – to different functions or to different second level properties.

If “the phi” and “the psi” are read as terms that refer to individuals, “…the phi…” and “…the psi…” bear the very same proposition. If “the phi” and

"the psi" are read any other way, they do not corefer (they either don't refer at all or refer to different things).

A difference in sense is thus *not* the consequence of replacing referring terms with co-referring terms. There are times when such a replacement *seems* to result in a difference in sense. But what is in fact going on is that, in such cases, the so-called "referring term" is functioning as a quantifier, and the difference in sense involves the replacement of a term with a term that does *not* co-refer.

Thus replacing a referring term with a co-referring term never results in a change of sense; SS is false. And the Frege-Church argument is destroyed.

By itself, this doesn't mean that sentences *do not* refer to truth-values. It means only that the Frege-Church argument for that thesis fails.

Token-semantics and the slingshot The question we are trying to answer is: what do sentences refer to? So far we've allowed that the question makes sense and embodies no confusions.

But given some pretty uncontroversial points, it is overwhelmingly likely that it is sentence-*tokens*, not sentence-*types*, that refer.

So far we've operated on (what I believe to be) the fiction that sentences *simpliciter* refer. Relative to that fiction, we've made a case that sentences do not refer to truth-values.

I will now produce an argument purporting to show that sentence-*types* do not refer any more than the expression-types "you" or "today" refer. The issue is not "what do sentence-types refer to?" It can be shown, on grounds acceptable to everybody -- both friends and foes of the slingshot -- that sentence-*types* don't refer to anything, and that the real question is: What do sentence-*tokens* refer to?

So far I've operated on the simplifying assumption that sentences

simpliciter refer: I haven't separated the problem of sentence-*token*-reference from that of sentence-*type*-reference.

Within the scope of that simplifying assumption, we've seen that sentences probably do not refer to truth-values.

It now time to drop that simplifying assumption.

The distinction between sentence-tokens and sentence-types plays virtually no role in discussions of the slingshot. Neale, Church, Frege, Quine, Davidson, and Gödel never make the distinction between sentence-meaning and sentence-type meaning.

They have all presupposed a "type-semantics": it is sentences *per se* that have propositions for their literal meanings. It is the sentence "snow is white" *per se* that bears the proposition *snow is white*.

As we've seen, given a type-semantics, the slingshot fails. But a compelling case for "token-semantics" is not hard to make. (By "token-semantics" I mean the doctrine that it is *tokens* of sentences that bear propositions, not sentence-types.) If we are to be precise and thorough in our treatment of the slingshot, we must deal with the possibility that token-semantics is correct; and re-express our arguments within that framework. It is not hard to do, and will take only a moment.

sentence-type reference versus sentence-token reference

We've seen that, in at least some cases, the meaning of expression-*types* must be distinguished from the meaning of expression-*tokens.* Of course, this is not a new discovery. Strawson (1951) was extremely clear about it; and Kaplan (1989/1977) used it to great advantage in his work in indexicals.

But the significance of this distinction has not been fully appreciated, at least not in connection with the question whether sentences refer.

Consider the sentence-*type* “I am tired”. (What follows is really due to Kaplan 1989.) This sentence-type has a meaning. But that meaning is not a proposition. The meaning is a rule that assigns propositions to *tokens* of that sentence, on the basis of the context of utterance. So the meaning of the type “I am tired” is some rule; it is not a proposition.

Where indexical-sentences are concerned, the context is defined by many things – different things, depending on which indexical is involved. But if somebody says (for example) “I am tired”, the context of utterance includes the world of utterance, the time of utterance, and the person who produced the utterance.

The intuitive analysis of definite descriptions is this: “the phi” picks out the phi (if there is one). “The inventor of bifocals” picks out the inventor of bifocals. We’ve seen that, if this is right, it is not the expression-type that refers, but tokens of it. The meaning of the expression-type is a rule that assigns a referent to tokens of it depending on the context of utterance.

So the meaning of “the inventor of bifocals snored” is not a proposition. It is a function from contexts of utterance to propositions.

The meanings of sentence-*types* containing indexicals and (possibly) definite descriptions are functions, not propositions.

It is clear that *some* sentences have a two-dimensional structure, meaning that it is *tokens* of them, not the types themselves, that bear propositions. It seems to me that it would be extremely arbitrary to suppose that *some* sentences have such a structure, and others do not. What is *patently* going on in the case of indexical-sentences, and sentences with definite descriptions, is *always* going. Actually, once it is realized that tense markers are indexicals, it turns out to be very hard to produce a sentence that does not have an indexical component[207], and thus a two-dimensional semantics.

In natural language, it is not possible to produce such a sentence; we have

to use an artificial notation to do so. But, of course, in some cases, the tense-marking is obviously irrelevant. If I say “2 +2 is 4”, it is patently obvious that the “is” isn’t doing any work. But that doesn’t mean that *semantically* the “is” is not an indexical. We know, on independent grounds, that considerations of time-indexing are irrelevant in this kind of case; so we ignore the tense-marker – we ignore the indexical. But surely at the level of *semantics* the “tense marker ” in “2 plus 2 is/equals 4” is the same as the tense-marker in “John is in Ohio”. It would be very strange to suppose that we were dealing with semantic ambiguity here. We could use Kripke’s classic (1991/1977) argument to show that.

That argument, adapted, is: If it *were* a case of semantic ambiguity, we’d expect there to be languages that had *different* words or morphemes: one for the innocuous tense-marker in “2 plus 2 is/equals 4”, and another for the non-innocuous one in “John is in the kitchen”. I doubt there are such languages. So we are not dealing with semantic ambiguity.

So *semantically* all natural languages sentences are indexical-sentences. Thus sentence-*types* don’t have propositions for the meanings; only sentence-tokens do – even when we’re dealing with *supposedly* eternal sentences. (Natural language cannot, at the level of semantics, produce truly eternal sentences. Whenever philosophers *say* that some sentence is eternal, they always make an ad hoc stipulation to the effect that the tense-marker/indexical is “timeless”. But that stipulation amounts to: don’t give the “-s” the meaning that it *actually* has.)

So a natural analysis is to this. Sentence-types *per se* don’t bear propositions. They bear functions from contexts of utterance to propositions. Sentence-*tokens* bear propositions. If there are any genuinely eternal sentence-types, then *their* meanings are functions: they just happen to be constant functions.

So the question “what do sentences refer to?” is ambiguous? The semantic content of a sentence-*type* is a function. The sentence-type doesn’t refer to anything, just as the expression-type “you” doesn’t refer to anything.

Frege and Church would admit that the sentence “Socrates was wise” has different truth-values, and thus refers to different things, in different possible worlds. (They didn’t talk in terms of possible worlds. But that doesn’t really affect the point.) So, if what I argued earlier is cogent, then they would be forced to go two-dimensional: so different tokens of that sentence have different semantic values; the type is invariant with respect to those changes in truth-value. Frege and Church would grant that “Socrates was wise”, with its actual semantics, could have different truth-values, and thus different referents, in different worlds.

So they must go two-dimensional; they must say that *tokens* of that sentence refer, and that the sentence-type doesn’t refer at all. Otherwise they are forced into the absurd position that the “Socrates was wise” can have the same semantic content that it has in this world, and yet have a different referent. So they must say that the referent of “Socrates was wise” is not part of its semantic content. But this is absurd.

We must say this: The type “Socrates was wise” doesn’t refer to anything, just as the expression-type “you” doesn’t refer to anything. Sentence-tokens, not sentence-types, refer.

The real question is: what do sentence-*tokens* refer to? And the live options are two. (i) Sentence-*tokens* refer to truth-values. (ii) Sentence-*tokens* refer to propositions.

Chapter 10: token-semantics and the slingshot

Let us very briefly, omitting details, re-express what we've said in a token-semantics framework.

(i) Sir Walter Scott is Sir Walter Scott.
(ii) Sir Walter Scott is the author of Waverly.
(iii) Sir Walter Scott is the man who wrote twenty-nine Waverly novels altogether.
(iv) The number n, such that Sir Walter Scott wrote n Waverly novels, is twenty-nine.
(v) The number of counties in Utah is twenty-nine.

Let (i)-(v) be sentence-*tokens* (in this world).

We can read the descriptions either referentially or attributively.

Read referentially, they are synonymous with:

(i_R) Sir Walter Scott is Sir Walter Scott.
(ii_R) Sir Walter Scott is Sir Walter Scott.
(iii_R) Sir Walter Scott is Sir Walter Scott.
(iv_R) Twenty-nine is twenty-nine.
(v_R) Twenty-nine is twenty-nine.

Assume, of course, that these are sentence-tokens (in this world).

This sequence of sentences obviously doesn't show that any two true sentence-tokens co-refer. (CR) can no longer guarantee the needed steps.

This is not a peculiarity of the sentence-tokens involved in (i)-(v). Choose any sentences you wish; run an analogue of Church's argument. Read

referentially, the resulting sequence is as innocuous as (i_R)-(v_R).

So let us give an attributive reading to (i)-(v):

(i_A) Sir Walter Scott is Sir Walter Scott.
(ii_A) Somebody x uniquely wrote Waverly, and Sir Walter is identical with x.
(iii_A) Some man uniquely wrote twenty-nine Waverly novels altogether, and Sir Walter Scott is identical with x.
(iv_A Some number n is uniquely such that Scott wrote n Waverly novels altogether, and n is twenty-nine.
(v_A) Some number n is uniquely such that there are n counties in Utah, and n is twenty-nine.

This sequence of sentence-tokens doesn't show anything. The last two installments are not even remotely close to being synonymous, contrary to what Church's argument requires.

Exactly similar considerations apply to Quine's and Davidson's arguments, after they are recast as sequences of sentence-*tokens*. It would be tedious and necessary to show this.

It is clear that the validity of our main argument against the slingshot is independent of the question whether type-semantics or token-semantics is right approach.

Of course, our other arguments apply are equally unaffected. Obviously one can know the semantic value of a sentence-*token* without knowing its truth-value. So it is hard to see how truth-value could be semantically contributed.

The only way to make the slingshot invulnerable to this last criticism is to say: the referent of a term is not a part of its semantic content. But that very line is a by-product of a *failure* to register the considerations that favor token semantics. The old view that a given expression can have *both* meaning and reference is a confusion, deriving from a failure to distinguish expression-types from expression-tokens – a failure to distinguish the *type* "you" from its tokens. Tokens refer; types do not. And once that distinction is made, it makes perilously little sense to say of something that refers that one can know its semantic content without knowing its referent.

Also, if we recast Slingshot-argument as sequences of sentence-tokens, we get the absurd results discussed earlier. We get Quine's intended but absurd view that necessity collapses into actuality; Davidson's intended but absurd view that there is only one fact; and unintended and absurd view that there is only proposition. The token-Slingshot creates these monsters no less than the type-Slingshot.

Reference re-analyzed

We analyzed reference thus: E refer to O iff sentences of the form "… E…" mean:…*O*…

Given that token-semantics is correct, we must change this to:

Tokens of E refer to O iff sentence-tokens of the form "…E…" mean:… *O*…

This conception of reference, it is easy to verify, does all the anti-Slingshot work that its type-semantics counterpart did.

Part II. a positive account of sentence(-token)-reference

Chapter 11. the meanings of sentence-types

sentence-type meaning and reference

What is the meaning of a sentence-*type*?

The old answer – the type-semantics answer – is: it is a function from worlds to truth-values.

But this is false. The meaning of the type "I am tired" is a function from contexts of utterance to propositions.

Those propositions *in their turn* may be functions from worlds to truth-values. But the sentence-type "I am tired" is surely not such a function; for it isn't true or false at all; so it doesn't assign truth or falsity to worlds or anything else.

We could say that the meanings of *non*-indexical sentences-types are propositions. But (a) natural language contains few, if any, such sentences; and (b) it would be very strange if the meanings of *some* sentence-types were functions from contexts to propositions, while the meanings of other sentence-types were functions from worlds to truth-values.

The sentence-*type* "I am tired" is the same sort of thing as the sentence type "2+2=4". But if the meaning of the one were a proposition, while the other meaning of the other were a function from contexts to propositions, then we'd be dealing with things of entirely different semantic orders. Under that circumstance, it would be extremely misleading to describe them both as *sentence*-types. "Sentence" is supposed to a semantic term; when you say that such and such is a "sentence", or a "sentence-type", you are saying that it has

a certain kind of thing for its meaning. If the meanings of the "sentence"-types "2+2=4" and "Bob snores" were entirely different things – if one was a function from worlds to truth-values, while the other was a function from contexts to propositions (i.e. from contexts to functions from worlds to truth-values) – then semantically those two entities would be too different to legitimately call by the same name.

The natural view is to say that, uniformly, the meaning of a sentence-type is a function from contexts to propositions, *not* from worlds to truth-values.

So the debate between those who believe that "sentences" refer to truth-values and those who deny that is really a debate as to what sentence *tokens* refer to.

Sentence-*types* don't refer to anything, any more than the *type* "today"; they don't bear propositions; they don't have truth-values; they don't refer to such things. Sentence-tokens are what refer.

summary

Uncontroversially, indexical sentence-types don't bear propositions; they bear functions from contexts to propositions. Tokens of such sentences bear propositions. It would introduce an artificial cleavage into the class of sentences if we said that some-sentence types had propositions for the meanings while others had functions from contexts to propositions. So we must assume that all sentence-types bear such functions, and that it is only sentence-tokens that have propositions for the meanings. In any case, all sentences of natural language are indexical sentences. So there is no escaping token-semantics. So the question is not "what do sentences refer to?", but

“what do sentence-tokens refer to?”

Chapter 12 *What sentence-tokens refer to*

A terminological point

Henceforth, until the very last chapter, everything I am going to say will concern sentence-*tokens*, not sentence-*types.* If I were to use the word “token” in every place it was required, the result would be intolerable verbosity and phraseological awkwardness.

So henceforth, whenever I talk about “sentences”, I really mean sentence-*tokens.* And whenever I talk about “the sentence ‘snow is white’”, I really mean: *tokens* of the sentence “snow is white”. (When I say, for example, “sentences refer to proposition”, I mean: sentence-*tokens* refer to propositions. When I say “snow is white” refers to a certain proposition, I mean: *tokens* of “snow is white” refer to that proposition.

our positive analysis

Sentences refer to propositions. Just as “Socrates” refers to Socrates, so “snow is white” refers to *snow is white.*

Remember what we said about reference. What is it for “Socrates” to refer to Socrates? It is Socrates to be the semantic contribution of “Socrates” to the propositions borne by sentences in which that expression occurs. Put another way, it is for

"…Socrates…"

to mean

…Socrates….

If "Socrates" contributed *Plato* to sentences in which that expression occurred, then "Socrates" would refer to Plato. So if a term is to refer to Socrates, it is *necessary* that its semantic contribution be Socrates.

If "Plato" (or any other term) contributes *Socrates* to sentences in which that expression occurred, then "Plato" would refer to Socrates. So if a term is to refer to Socrates, it is *sufficient* that its semantic contribution be Socrates.

If y is the semantic contribution of x, then x *refers* to y. This is a consequence of what we said a moment ago.

Right now I will give a rough and ready argument to show that sentences refer to propositions – the very propositions they express or bear. I will deal with the niceties after putting forth the bare bones of the argument.

our argument (an outline)

Consider the sentence "Bob smokes". What is the semantic contribution of that sentence? What is meant by "...Bob smokes…"?

"…Bob smokes…" means:…*Bob smokes.…*

The semantic contribution of "Bob smokes" is the proposition *Bob smokes.* (For reasons discussed earlier, the truth-value of that sentence is not among its semantic contributions.)

Let us consider *one* sentence of the form "…Bob smokes…". One such

sentence is:

(a) “Bob smokes.”

Of course, (a) means:

(b) *Bob smokes*.

So, in the case just considered, the semantic contribution of “Bob smokes” is: *Bob smokes.*

As it turns out, the semantic contribution of “Bob smokes” is *always* the proposition: *Bob smokes.*

Consider the sentence:

(c) “Plato snored and Bob smokes”.

This means:

(d) *Plato snored and Bob smokes*.

Consider:

(e) “John believes that Bob smokes.”

This means:

(f) *John believes that Bob smokes.*

In each case, the semantic contribution of “Bob smokes” is: *Bob smokes.*

Any sentence of the form

“…Socrates…”

means (or is true if and only if)

…*Socrates.*…

And that is necessary and sufficient for “Socrates” being a term that refers to Socrates.

Any sentence of the form

“…Bob smokes…”

means (or is true if and only if)

…*Bob smokes.*…

It is starting to seem – though we haven’t proven it yet – that “Bob smokes” refers to the proposition: *Bob smokes.* Once it is granted that reference is semantic contribution, it is hard to see how we could say anything else about “Bob smokes”. If we deny that “Bob smokes” refers to that proposition, then by implication we are saying that Socrates could be the semantic value of “Socrates” *without* that expression’s referring to that

person.

Chapter 11. dealing with some objections

Now we must deal with some objections to our theory. These objections will require that we do some polishing. Here is the most important objection (I will call it the “true-false objection”). I will put this objection in the voice of a hostile interlocutor.

The true-false objection

It is obviously a matter of controversy whether “Bob smokes” refers to the proposition: *Bob smokes.*

But some expressions *uncontroversially* refer to that proposition. “That Bob smokes” is one such expression. As English speakers, when we want to refer to a proposition, we simply put a “that” in front of a sentence that bears that proposition. (Significantly, the “that” is often omitted, as in “Bob believes Frank smokes”.) Every language has an equivalent device.[208]

(a) “Bob smokes”

is true or false.

(b) “That Bob smokes”

is not true or false.

Here it is important to distinguish expressions from their referents. The

expression

(a) “that Bob smokes”

is neither true no false. (If I say “that Bob smokes”, I haven’t said anything true or false.) But the thing that (a) *refers* to is true or false.

(c) *That Bob smokes*

is true or false. It is a proposition, and propositions are always true or false.

In general, referring terms are not true or false (it seems). “Bob” is not true or false. And this precludes sentences from referring at all, which nullifies your attempt to show that they refer to propositions.

So there are two problems for your theory. First, it seems that referring terms are not true or false. If true, this point devastates our theory. For it would mean that sentences, being true or false, cannot be referring terms.

But even if some referring terms are true or false, your theory *still* has a problem. If (a) and (b) are really co-referential, then it seems strange that one of them should be true or false, while the other is neither. Given that (a) is true or false, while (b) is neither, it is *de rigueur* to suppose that there is some *semantic* difference between them.

Both (a) and (b) are associated with the same proposition. So it would be hard to argue that they differ in *sense* or in *intension.* It cannot be denied that (b) *refers* to the proposition *Bob smokes.* So we are blocked from saying that (a) and (b) differ in sense or intension. And we are blocked from denying that (b) refers to *Bob smokes.* But we must account for the obvious difference between (a) and (b). One makes a statement; the other does not. So, presumably, there is some kind of *semantic* difference between them.

The only way to go, it seems, is to say that (a) does *not* refer to anything.

This exposes a fallacy in your reasoning. You say that the semantic contribution of “Bob smokes” is the proposition: *Bob smokes.* Obviously this *seems* true. For obviously “Bob smokes” means *Bob smokes*.

But the semantic contribution of “Bob smokes” is *not* in fact: *Bob smokes.* Again, consider the sentence “Bob smokes.” This is certainly *a* sentence of the form “…Bob smokes…”. But “Bob smokes” does not mean: *that Bob smokes*. It means: *Bob smokes.* If it meant the former, it wouldn’t be true or false.

Obviously “that Bob smokes” means *that Bob smokes.* That proposition is the semantic contribution of that expression. But “that Bob smokes” is not true or false. What it means isn’t true or false. But what “Bob smokes” means *is* true or false.

So “Bob smokes” does *not* contribute the proposition *that Bob smokes* to the senses of the sentences in which it occurs. If that *were* its contribution, it wouldn’t be true or false. But it is true or false.

dealing with the true-false objection

The true-false objection is based on a failure to appreciate the concept of sentence-*force.*

Consider the sentence

(a) “Bob smokes”

and

(d) “does Bob smoke?”.

The first is true or false; the second is neither – it is no more capable of truth or falsity than “that Bob smokes” or “Socrates”.

But (a) and (d) seem, at least arguably, to bear the same proposition: they have the same *sense* or *intension* (on any delineation of these terms).

We don’t want to say that they differ in referent or *extension.* Given that they have the same intension or sense, they cannot possibly *differ* in referent or extension. Surely the *intension* of a sentence fixes what its *extension* is. “Schnee is weiss” and “snow is white” have the same intension; and it thus would be madness to say that have different extensions. “Bob smokes” and “does Bob smoke?” surely have the same intension; so, by parity of reasoning, we must say that they cannot differ in extension.

But (a) and (d) do differ in *some* semantically significant respect. So intension and extension, sense and reference, are not *all* there is to semantics – at least not where the semantics of *sentences* are concerned. There is a third-component to sentence-semantics: *force.*

(a) and (d) have different *forces.* (a) has indicative force; (b) has interrogative force.

In light of this point, let us re-consider

(a) “Bob smokes”

and

(b) “That Bob smokes”.

It is true that (a) is true or false, while (b) is neither. But *by itself* this doesn't mean that (a) and (b) differ in referent *or* that they differ in sense. For (a) and (e) cannot possibly differ in referent or in sense. And yet one of them is true or false, while the other is neither.

So a difference in force cannot possibly constitute a difference in sense (intension) or a difference in reference (extension).

This is not to say that sentences *do* have referents or extensions. But it is to say that *if* they do, then differences in sentence-forces do not indicate difference in sentences-referents (-extensions).

This is the first step in our response to the true-false objection.

Here is the next step. "That Bob smokes" does not have force. What it refers to is true or false. But that expression *itself* is not true or false: for it lacks assertoric force. It must be given assertoric force. *Then* it will be true or false.

Let us imagine a language – call it English* -- that is just like English except in the following respect. (For the moment, let us focus on written English*.) In English*, when one wants to *assert* that Bob smokes, one doesn't say

(a) "Bob smokes".

Instead, one says:

(f) "It is true: that Bob smokes."

And when one wants to know whether Bob smokes, one doesn't say:

(g) Does Bob smoke?

Instead, one says:

(h) “Is it true: that Bob smokes.”

Thus the English*-translation of the English sentence “Bob smokes” is (f). The English*-translation of the English “does Bob smoke?” is (h). And so on.

Is English* a *different* language from English? English* is a *fragment* of English. English has *two* ways of saying that Bob smokes: “Bob smokes” and “It is true: *that* Bob smokes.”

English* has only one way of saying this. Otherwise English* is just like English.

“It is true: *that Bob smokes*” and “Bob smokes” are synonymous. In English, given an expression referring to a proposition, one of way of giving it assertoric force is to drop the “that”. Dropping that “that” from “that Bob smokes” is the same thing as prefixing an “it is true” to it. The only differences are phonological or orthographic.

To make “that Bob smokes” be true or false, all we have to do is give it a *force* – assertoric force. To make it express a sentence or a question, we must give it some other force.

“Bob smokes” and “that Bob smokes” differ only in *force.* One has assertoric force; the other has no force.

A difference in *force* cannot possibly constitute a difference in *sense* or in *reference.* We saw this earlier. (I will give another argument in a moment.)

Uncontroversially, “That Bob smokes” refers to the proposition *that Bob smokes.*

Thus “Bob smokes” *refers* to that same proposition.

summary

Uncontroversially, “that Bob smokes” refers to the proposition: *that Bob smokes.* “….that Bob smokes…” means:…*that Bob smokes.*..That is what it is for “that Bob smokes” to refer to that proposition. That expression contributes that proposition to the propositions borne by the sentences in which that expression occurs. There is no difference between contributing that proposition in such contexts, on the one hand, and referring to that proposition, on the other.

Let us now switch from talking about “that Bob smokes” to talking about “Bob smokes”.

Uncontroversially, *part* of what “Bob smokes” contributes is the proposition: *that Bob smokes.* If you say “Bob smokes”, surely *part* of the meaning borne by your words is the proposition: *that Bob smokes.* The same is true of any sentence of the form “…Bob smokes…”

Thus, uncontroversially, at least *part* of what “Bob smokes” does is to contribute *that Bob smokes.* So at least *part* of what “Bob smokes” does is to refer to: *that Bob smokes.*

There is a difference between “that Bob smokes” and “Bob smokes”. The second is true or false; the first is neither.

But that difference is *not* due to their contributing different propositions.

Thus, uncontroversially, the *difference* between “Bob smokes” and “that Bob smokes” does not lie a difference in what they refer to. They refer to the same thing: the proposition *that Bob smokes.*

Changing or adding force to an expression referring to a proposition cannot change its referent: another argument

The argument just given relies on the following principle:

If an expression *lacks* a force, giving it one cannot change its referent. And if it *has* a force, giving it a new one cannot change its referent.

This, I think, is the *only* step in my argument that is debatable. But it is still hardly debatable. Let me now give a slightly more developed defense of it than was given before.

"Bob smokes" is true or false, while "that Bob smokes" is neither. That must be granted.

Uncontroversially, the semantic contribution of "that Bob smokes" is the proposition *that Bob smokes.* It is also uncontroversial that at least *part* of the semantic contribution of "Bob smokes" is the proposition *that Bob smokes.* Surely a *proposition* is, at the very least, *among* the things meant by "Bob smokes". No one denies that. And surely the proposition in question is: *that Bob smokes.* No one denies *that.*

So it is undeniable that, at least *up to a point*, "Bob smokes" and "that Bob smokes" both mean *that Bob smokes.* Both expressions contribute that proposition.

How do the semantic contributions of those expressions *differ*? Obviously they don't differ in respect of the *proposition* that they contribute. It is not as though one of those expression contributes a proposition that the other does not.

Admittedly, this last point might be questioned. One might say this:

"Bob smokes" contributes the proposition: *it is true that Bob smokes,* while "that Bob smokes" merely contributes the proposition: *that Bob smokes.* And *this* is why "that Bob smokes" is neither true nor false, while "Bob smokes"

is true or false. A difference in force *does* reduce to difference in the proposition contributed.

But that is absurd. Even if we grant that “Bob smokes” contributes the proposition *it is true that Bob smokes*, as opposed to (merely) *Bob smokes*, that has nothing to do with the difference in *force* between “Bob smokes” and “that Bob smokes”. For the expression “that it is true that Bob smokes” is neither true nor false. But, uncontroversially, the semantic contribution of that expression is the proposition: *that it is true that Bob smokes.* Differences in *force* are not reducible to differences in propositional content.

So supposing that “Bob smokes” contributes the proposition *that it is true that Bob smokes* does absolutely nothing to explain why “Bob smokes” is true or false, while “that Bob smoke” is neither.

So it is undeniable that “Bob smokes” and “that Bob smokes” *both* contribute the proposition *that Bob smokes.* It is also undeniable that the difference between “Bob smokes” and “that Bob smokes” is *not* due to a difference in the propositions that they contribute. In general, changing the *force* of an expression does not change its semantic contribution. Giving force to an expression that lacks it won’t change its force; and neither will change the force of an expression that already has it.

Chapter 12. cleaning up Frege’s mess

Our analysis enables us to clean up a mess that has been sullying the logico-philosophical landscape for over a century. In the process of cleaning up this mess, we will clarify important features of our own positive view.

In some contexts, replacing a sentence S with another sentence of the same truth-value preserves truth-value. Consider the sentence:

(k) “grass is green and snow is white”

Replacing “grass is green” with another sentence of the same truth-value results in a sentence that has the same truth-value as (k):

(l) “Frege defined numbers as sets of sets and snow is white”.

A limiting case of such a context is a single sentence. If we replace

(m) “snow is white”

with a sentence of the same truth-value, what results has the same truth-value as the original.

Obviously not all contexts are like this.

(n) “Little Timmy believes that grass is green”.

If we replace “grass is green” with a sentence of the same truth-value, what results doesn’t always have the same truth-value as the original:

(o) “Little Timmy believes that Frege defined numbers as sets of sets”.

Frege concluded from this that “grass is green” has one meaning when it occurs by itself, or in “grass is green and snow is white”, and a completely different meaning in “Little Timmy believes that grass is green.”

So “grass is green” means one thing in:

(p) “grass is green”

And it means something else in (n).

Intuitively, this seems very wrong. Surely “grass is green” has the same meaning in both cases. (My own view, of course, is that “grass is green” *always* refers to the proposition: *that grass is green.* It doesn’t matter what kind of context it occurs in.)

There is an extremely interesting objection (due to Nathan Salmon) that we must consider:

You very reasonably say that “grass is green” has the same meaning in both

(p) “grass is green”

and also

(n). “Little Timmy believes that grass is green.”

But this is perhaps consistent with its referring to a truth-value in (p) but not in (n). Roughly, it may function as a Millian name of the proposition in (n), and as a description of a truth-value in (p). If I am right about this, then we can hold onto the idea that “grass is green” has the same meaning, at least up to a point, in both (p) and (n), while also holding that it refers to a truth-value (in extensional contexts).

I'm not entirely sure I understand this objection. But what I think it means is this. Consider some definite description, say "the teacher of Aristotle". This description can function either referentially or attributively. If you say "Little Timmy belies that the teacher of Aristotle was wise", you may be saying that Little Timmy believes something about *Plato* (referential reading) or you may be saying that Little Timmy believes that *whoever it was* that taught Plato was wise (attributive reading).

There are very good reasons for thinking that "the teacher of Aristotle" is not semantically ambiguous, and that is amenability to both referential and attributive *readings* is a matter of pragmatics, not semantics.

The objector seems to be saying that "grass is green" is, in this respect, like a definite description. In (p) it refers to a truth-value (it refers to it by encoding descriptive information that singles it out). So in that context it is functioning "referentially". But in (n), "grass is green" refers to a proposition. So it is functioning "attributively".

But even though "grass is green" may *function* in both ways, it doesn't follow that it is semantically ambiguous. The ambiguity might be completely pragmatic, like that of "the teacher of Aristotle".

So maybe "grass is green" *does* refer to truth-values; the fact that sometimes *seems* not to do so is some kind of pragmatic epiphenomenon.

Some support (here I may or may not be going beyond what the objector says) lies in the fact that, arguably, "that grass is green" *is* a definite description ("the proposition that grass is green").

Let us take stock.

First of all, this is about the best proposal I've ever encountered for taking the sting out of the Fregean view that "grass is green" is ambiguous between truth-values and senses. Indeed, I would go so far as to say that one *must* take this view, if one is to hold that sentences refer to truth-values. Otherwise one

must hold the unpalatable view that “grass is green” is infinitely ambiguous.

Nonetheless I have some reservations.

First of all, if “(that) grass is green” describes anything, it probably isn’t a truth-value. “that grass is green” is, as we noted, nearly synonymous with “the proposition that grass is green”. And the latter describes/refers to a proposition, not a truth-value.[209]

But let us waive this; let us suppose that “grass is green” does describe a truth-value.

Here is the problem (we’ve encountered it before). Obviously “grass is green” has different truth-values in different worlds. (And even if it doesn’t, just replace “grass is green” with a sentence that *does* thus vary in truth-value.) Let w be this world, and let w* be some world where English is spoken but grass is purple.

If the objector is right, then in w, “grass is green” describes the truth-value *true*. In w*, it describes the truth-value *false*.

Since English is spoken in w*, “grass is green” has the same semantics there that it has here in w.

Obviously the referent of an expression is at least *part* of its semantic content.

Therefore, in w*, “grass is green” doesn’t have (quite) the same semantic content that it has here in w.

But obviously “grass is green” has the same semantic content in both cases. A dilemma.

We’ve dealt with this sort of situation before. The natural way (the *only* way, probably) of dealing with it would be to say:

“Grass is green” has a two-dimensional semantic structure. In a world where grass is green, *tokens* of it refer to (by describing) the truth-value *true*. In a

world where grass is not green, tokens of it refer to the truth-value *false.*

So “grass is green” is a non-rigid designator; and whether a token of it picks out the true or the false is a function of the context (world) of utterance.

If this is right, then it is not “grass is green” *per se* that picks out a truth-value, but rather tokens of it. And the expression “grass is green” *per se* wouldn’t refer to anything; its meaning would be a function that picked out truth-values on the basis of the context (world) of utterance.

A consequence is that the semantic content of a *token* of “grass is green” is a truth-value, while the semantic content of the expression-type is the just mentioned function.

Now we can see a problem with the objector’s approach. If he is right, then one cannot know the semantic content of a token of “grass is green” unless one knows its truth-value in the context (world) of utterance (just as one cannot know the semantic content of an indexical unless one knows what it refers to in the context of utterance). In general, one cannot know the semantic content of a token of a sentence unless one knows its truth-value in the context (world) of utterance

But it is hard to believe that I don’t know the semantic content of a token of a sentence if I don’t know its truth-value. To understand a sentence-token – to understand a token of “that guy is tired” – is to know what *would* have to be the case *if* it were true; it is not to know *whether* it is true. The most natural position, I think, is to say that one *can* understand sentence-tokens without knowing their truth-values; and, consequently, that such tokens do not refer to truth-values.

And, as we’ve seen, it is not easy to make the case that sentence-*types* refer to truth-values. For obviously the truth-value associated with the sentence-*type* “grass is green” varies from world to world, even though its

semantics does not (or need not) so vary. So unless we are willing to take the heroic measure of saying that the referent of an expression is not part of its semantic content, we should probably distance ourselves from the idea that sentence-types refer to truth-values.

Chapter 13. dealing with Frege's view

Frege *did* make a point about sentence-reference that, I think, is quite correct but that has been given its due. He said that, in (n), "grass is green" refers to the proposition *grass is green.*

He put it this way. In "oblique contexts" sentences refer to the propositions that, in non-oblique contexts, are their *senses.* (n) is an oblique context. So "grass is green" refers to the proposition that, ordinarily, is its sense, namely: *that grass is green.*

But Frege *also* says that when "grass is green" occurs on its own, or in the scope of a truth-functional connective, it does *not* refer to the proposition *that grass is green.* This is where he is wrong. And it is this false belief that leads Frege to make his exceedingly strange claim that "grass is green" has altogether different meanings in "grass is green" (occurring on its own) and "Little Timmy believes that grass is green".

Intuitively we want to say that "grass is green" means the same thing in both (n) and in "grass is green" (occurring on its own). At the same time, we must acknowledge that, whereas the *truth* of "grass is green" (occurring on its own) depends on the *truth* of the proposition *that grass is green*, the truth of (n) does not so depend. (n) could be true even if grass were orange.

Here is what I would suggest. The difference lies *not* in what "grass is green" refers to. "Grass is green" refers to the same thing in both contexts.

The difference lies in the *force* that is attached to that expression. Remember that saying “grass is green” is the same thing as saying “that grass is green” with assertoric *force*. Dropping the “that” is simply a way of indicating that assertoric force is attached.[210]

In (n), “grass is green” has no force, and thus no assertoric force. What has assertoric force in (n) is the *whole* sentence, not “grass is green”. But when “grass is green” occurs on its own, what has assertoric *force* is, in effect, “that grass is green”.

Granted, this last claim may seem odd. For when one says “grass is green”, by itself, the expression “that grass is green” simply isn’t *there* to be given force. A proper part of that expression seems to be there, but not the expression itself.

This is a mistake. When you say “grass is green”, what you are doing is giving assertoric force to the expression “that grass is green”; and the device by which you do this is to drop the “that”. So when you say “grass is green”, the expression “that grass is green” *is* there, though not in its usual phonetic garb.

Any imperative contains a phonetically unrealized “you”. (If you say “march!”, you are really saying: “you: march!”.) If somebody asks you, “who was the greatest philosopher of antiquity?”, and you say “Socrates”, your answer is elliptical for “Socrates was the g greatest philosopher of antiquity”; so your answer contains a phonetically unrealized “was the greatest philosopher of antiquity”. Philosophy of language is impossible to do without countenancing phonetically unrealized components.

The proposition semantically contributed by “grass is green” is always *grass is green*; it doesn’t matter whether “grass is green” is occurring on its own or in (n). What differs is that, in the one case, assertoric force attaches, whereas in the other case, assertoric force does not attach.

Summary

The sentence:

(n) “Little Timmy believes that grass is green”.

refers to the proposition: *that Little Timmy believes that grass is green.* Of course, (n) is true or false, whereas the expression

(q) “that Little Timmy believes that grass is green”

is neither true nor false. But this is because, in (q), no assertoric force is attached to the expression in question, whereas in (n), assertoric force *is* attached to that very expression.

So Frege’s argument is broken. It is true that the truth of “grass is green”, on its own depends on that of *that grass is green*. It is also true that the truth of (q) does not depend on that of *that grass is green.* But that is *not* because “that grass is green” refers to different things in those contexts. It is because, on the contrary, they refer to the *same* thing (*that grass is green*), but are associated with different forces.

Chapter 14. continuing to clean up Frege’s mess: compositionality and indirect discourse

Frege rightly believed in the principle of compositionality: The meaning of a complex expression is a function of the meanings of its parts. He also rightly held a certain specialized version of the principle: The referent of any

referring term is a function of the referents of its parts.

Frege held (rightly) that (n) is a referring term. So he also held (rightly) that the referent of (n) is a function of the referents of its parts.

It is obvious that in (n), "grass is green" contributes – refers to – the proposition *that grass is green.* Frege agrees. When "grass is green" occurs in a sentential context (in limiting cases, it may *coincide* with that context), what it contributes to that context is the proposition: *grass is green.*

Frege wants to hold onto the principle of compositionality. So he thinks (rightly) that the referent of (k) is a function of the referents of its parts.

But he thinks that the referent of (k) is a function of the *truth-values* of its components sentences, not of the *propositions* they bear. (He was led to believe this by his mistaken belief in the correctness of SS.) At the same time, Frege wants to hold onto the idea that what (k) refers to is entirely a function of the referents of its components. Thus Frege is forced to say that, in (k), "grass is green" contributes a *truth*-value, whereas in (n) "grass is green" contributes a *proposition.* Thus to hold onto compositionality, Frege is forced to say that "grass is green" is *ambiguous*, not merely like the word "bank", but in a more pervasive way.

Our analysis does not have this disastrous consequence. On our analysis, "grass is green" makes the exact same contribution to both (n) and (k). There is no shift in meaning, no ambiguity.

On our analysis, "grass is green" always refers to the proposition: *that grass is green.* So, in (i), "grass is green" refers to *that grass is green.* And, in (i), "snow is white" refers to *that snow is white.*

On our analysis, (k) as a whole refers to: *that grass is green and snow is white.*

On our analysis, (k) is coreferential with "that grass is green and snow is white". The only difference is that assertoric force attaches to (k).

I wish to be very clear on one point. According to my analysis, “grass is green” *always* refers to the proposition *that grass is green*. (The same is true *mutatis mutandis* for any sentence.) So in the expression “that grass is green and snow is white”, “grass is green” refers to *that grass is green.* And in

(k) “grass is green and snow is white”

“grass is green” refers to *that grass is green.* In any case, this is what our analysis says.

Our analysis is controversial. But I think it follows by uncontroversial steps from uncontroversial premises. Let me now show this.

What is *not* controversial is this. The referent of

(r) “that grass is green and snow is white”

is the proposition

(s) *that grass is green and snow is white.*

So if our analysis is to prove correct, we must reconcile the obvious fact that (r) refers to (s) with *our* hypothesis that “grass is green” always refers to the proposition: *that grass is green*. And our reconciliation must respect the principle of compositionality.

This can be done. Consider the sequence: *<that grass is green, that snow is white>*. We can attribute properties to this sequence. One such property is this: *is such that both its members are true.* The resulting proposition is:

(t) *<that grass is green, that snow is white> is such that both its members are*

true.

(s) and (t) are, I submit, the same proposition. (Some will balk at this thesis. I will defend it in a moment.)

Let us produce a sentence that *uncontroversially* has (t) for its meaning:

(u) "<that grass is green, that snow is white> is such that both its members are true."

On our analysis, (u) refers to the proposition:

(v) that: *<that grass is green, that snow is white> is such that both its members are true.*

There is no denying the following. The expression "that grass is green" refers to the proposition: *that grass is green.*

There is no denying that the referent of (u) is a function of the referents of its constituents, these constituents being (inter alia) "that grass is green" and "that snow is white".

On our analysis (u) refers to (v).

On our analysis (k) refers to (s).

We seen (s) and (v) are the very same proposition. (At any rate, we've agreed on this provisionally.)

So, on our analysis, (u) and (k) refer to the very same thing.

There is no denying that, in (u), "that grass is green" refers to: *that grass is green.*

It follows (nearly enough) that, in (k), "grass is green" refers to the proposition: *that grass is green.*

And this is exactly what our analysis says it refers to. For the core of our analysis just *is* the assertion that "grass is green" *always* refers to the proposition: *that grass is green.* So our analysis has been vindicated.

This is not to mention that our analysis is in complete agreement with our pre-theoretic intuitions. It is very hard to deny that "that grass is green" and "grass is green" *both* contribute the proposition: *that grass is green.* No intuitively acceptable analysis would deny this. No such analysis would deny that "grass is green", occurring on its own or in a truth-functional context, contributes the proposition *that grass is green.* No analysis that harmonizes with our intuitions denies that "that grass is green" contributes that same proposition in epistemic contexts (and other supposedly deviant contexts) that it contributes in transparent contexts.

Our analysis has "grass is green" *always* contributing the proposition *that grass is green.* It doesn't shift meanings.

Further, our analysis respects compositionality.

So our analysis is on the right side of all these desiderata and intuitions.

Frege's analysis respects compositionality, but only by saying that all sentences are ambiguous. Actually, Frege's analysis requires that "grass is green" be *infinitely* ambiguous. If we grant, with Frege, that epistemic operators induce a shift in meaning in "grass is green", then we must hold that "grass is green" refers to one thing in "grass is green"; to another thing in "John believes that grass is green"; to a third thing in "Frank believes that John believes that grass is green"; and so on. So Frege's analysis makes "grass is green" be infinitely ambiguous. But surely "grass is green" has the same meaning in all these contexts; it isn't ambiguous at all.

We must hold onto the idea that sentences have the same meaning – make the same semantic contribution – in all contexts. Otherwise, inevitably, we will make them, not just ambiguous, but infinitely ambiguous. (Iteration of

the opacity-inducing operator will add ambiguity to ambiguity.) It is also a datum that in “John believes that grass is green”, “grass is green” refers to the proposition *that grass is green.* To avoid making “grass is green” be infinitely ambiguous (I’m skipping a few steps), we must make it refer to that proposition when it occurs on its own or in truth-functional contexts.

discharging an assumption

Earlier I made a strong claim. I said that

(s) *that grass is green and snow is white.*

is the very same proposition as:

(t) *<that grass is green, that snow is white> is such that both its members are true.*

Everyone will agree that (s) and (t) are logically equivalent. But for P and Q to be the same proposition, logical equivalence isn’t enough (*1+1=2*, and *arithmetic is incomplete*). Some kind of *cognitive* equivalence is also needed. And, it could be argued, (s) and (t) don’t have the needed cognitive equivalence.

This is not the case. In so far as it seems to be the case, that is simply a by-product of the fact that different strings of verbiage are involved. Obviously “schnee is weiss” might *seem* to bear a different proposition from “snow is white”. But in so far as it does, that is simply because one isn’t making the

needed transition from the verbiage, which *does* differ, to the proposition, which does not. I submit that (s) and (t) are no more *cognitively* distinct than are the propositions [sic] *that snow is white* and *dass schnee is weiss.*

Let me now argue for this bold claim.

Let English* be just like English, except that English* word order is a bit different.

(Let us focus on written English*.) In English*, one doesn't writes "grass is green and snow is white". Instead, one says: "And (grass is green, snow is white)".

Otherwise, English* is just like English.

What would an English*-grammarian (not logician or philosopher) say about "And (grass is green, snow is white)"? In other words, given somebody who reads logical form off of grammatical form, what would he say about that sentence? He would say that it attributes truth to a certain set or sequence of sentences (or, more accurately, to the propositions expressed by members of that set). So he would say that "and" picks out a *property* of sets or sequences of propositions – just as "red" picks out a property of individuals and "true" picks out a property of sentences.

Is our English* semanticist wrong? Is he being misled by surface structure? It is very hard to see how. English* seems to make perspicuous the predicative character of the English "and".

Recall a point made earlier. In the sentence "grass is green and snow is white" *both* constituent sentences fall within the scope of the "and". "And" is something that *applies* to a set consisting of two members. This fact – and it is a fact - is *obscured* by English word order. It is made perspicuous only by an artificial notation like: And (grass is green, snow is white). This is the notation used by English*. This gives us good reason to see English*-grammar as giving us the logic of the English "and".

Also, to say that "and" *applies* to the set (grass is green, snow is white) is scarcely different from saying that it attributes a property to that set. So, again, we are led back to the predicate character of "and".

(Exactly similar remarks support the view that *any* sentence-connective – "or", "for some x"…-- is a second-level predicate.)

Significantly, in "Polish" notation, one says: "And (grass is green, snow is white)", not "snow is white and grass is green".

The *logical* properties of "and" are made perspicuous in English*. It is English, not English*, that obscures these properties. English* word-order makes it very clear that "and" denotes a property of sets (or sequences) of sentences. And that property is nothing other than: *is such that both its members are true*.

Word order has no logical significance.[211] Apparently, the function of "and", when it joins whole sentences[212], is simply to denote a property of a sequence formed by those sentences. Significantly, in "Polish" notation, one says "And (snow is white, grass is green)".

I must deal with an objection to this:

As you say, the English*-translation of "grass is green and snow is white" is "and (grass is green, snow is white)". According to you, this shows that "and" (in both English and English*) denotes a property of sets of proposition.

That is *one* way to go. But here is another. "And" denotes a relation between the designate of sentences. So the "and" in "grass is green and snow is white" denotes a relation between either the propositions or truth-values of those sentences, depending on what we take sentence-reference to be.

I agree with this. But I don't think it contradicts what I said. What is the relation in question? Presumably it is the relation borne by P to Q just in case both are true. And this relation, it seems to me, is the very relation borne by P to Q just in case the set (P, Q) has the property of being such that both its members are true. So I would not say that the objector's point completely invalidates my analysis of "and".

Chapter 15. why sentences refer to propositions and not truth-values: another argument

I'd like to give another argument for the thesis that sentences do not refer to truth-values.

Sentences of the form "…Caesar…" are about Caesar. More exactly: sentences of the form: "…Caesar…" mean:….*Caesar*…

That, we have argued, is what it is for "Caesar" refers to Caesar.

Sentences of the form "…that snow is white…" mean:… *that snow is white*…

That, we have argued, is what it is for "that snow is white" to refer to the proposition: *that snow is white*.

Consider any sentence of the form "…snow is white…" (Notice the absence of the "that".) Examples are "grass is green and snow is white", "necessarily, snow is white", and simply "snow is white".

Surely all those sentences are about the proposition *snow is white.* Granted, all of those propositions are about many things other than proposition; they are about snow and whiteness and, perhaps, the relation of property-instantiation. But there can be no denying that, in virtue of

containing the sentence “snow is white”, a sentence is about the proposition that snow is white. There can be no doubt that sentences of the form “…snow is white…” are *ipso facto* about the proposition that snow is white. “snow is white” says that the proposition *that snow is white* is true. In any case, this is a plausible analysis of it. “Necessarily, snow is white” says that the proposition *that snow is white* is necessarily true.

This gives us *prima facie* reason to believe that “snow is white” refers to the proposition *that snow is white.*

We said that for “Caesar” to refer to Caesar just *is* for sentences of the form “…Caesar…” to mean:…*Caesar.*…

Now sentences of the form “…snow is white…” do *not* mean:…*the truth-value true.*…

Sentences of the form “…snow is white…” *do* mean:…*snow is white.*…, just as sentences of the form “…Caesar…” mean:…*Caesar.*…

“snow is white” semantically contributes *snow is white* to sentences in which it occurs, just as “Caesar” semantically contributes Caesar to sentences in which *it* occurs. That fact about “Caesar”, we have argued, is *identical* with its referring to Caesar. By parity of reasoning, we must say that “snow is white” refers to the very thing it means, namely *snow is white.*

Sentences of the form “…snow is white…” do not mean:…*the truth-value true.*…This is clear. Consider the sentence “snow is white”; this, of course, is *an* example of a sentence of the form “…snow is white…” Obviously the meaning of “snow is white” is not: *the truth-value true.*

There are expressions that have that meaning. Arguably, an example is: “the truth-value *true*.” A clear example is Kripke-Salmon’s “ze truth-value *true*”. But “snow is white” is not among such expressions.

So if we accept the idea that, for E to refer to O is for E to semantically contribute O, then we must reject the idea that “snow is white” refers to a

truth-value.

Suppose you actually want to make a statement about the truth-value *true*. What you'll do is use some expression like: "the truth-value *true*." What you *won't* do is use "snow is white" or "grass is green".

Of course, there is an obvious reply:

If I say "grass is green", I *am* saying something about the truth-value *true*. I am saying that *grass is green* has that truth-value. So I *can* use "grass is green" to make a statement about that truth-value.

For the sake of argument, suppose "snow is white" refers to the truth-value *true*, just like "the truth-value true" (or "ze truth-value true"). In that case, presumably, "snow is white" ought to be able to perform the functions performed by "the truth-value true". But this isn't the case. Consider the sentence:

(lsp) "The last sentence Plato uttered had the truth-value *true*"

If you replace "the truth-value *true*" with "snow is white" or "grass is green", what results is rubbish:

(lsp*) "The last sentence Plato uttered had snow is white".

Of course, one can do some syntax-chopping to make the substitution work; one can say that (pt) is equivalent to (say):

(lsp**) "The last sentence Plato uttered had the same truth-value as *snow is white*."

But clearly the idea that “snow is white” refers to the truth-value *true* is under serious strain. If “snow is white really refers to the truth-value *true*, why is it so hard to it to make it act as though it did? Why is it so hard to get to it behave like expressions that we *know*, for certain, to refer to that truth-value?

the ambiguity problem

It is very hard to believe that “snow is white” makes one semantic contribution when it occurs on its own, or in contexts like “snow is white and grass is green”, and makes a different semantic contribution in “possibly, snow is white” and “not necessarily, snow is white”.

In the sentence “possibly, snow is white”, the sentence “snow is white” is not semantically contributing its truth-value. In that sentence, what is being discussed is the *proposition* that snow is white; what is being discussed is *not* the truth-value *true* or the truth-value false. So, unless we are willing to take the highly counter-intuitive view that “snow is white” undergoes some kind of semantic shift when “possibly” is put in front of it, we must say that it doesn’t refer to a truth-value even in truth-functional contexts.

A different sentence might help. Consider the sentence “Socrates did not drink hemlock”. The sentence “possibly, Socrates did not drink hemlock” is true. Presumably “Socrates did not drink hemlock” has the same meaning in both of those sentence; it doesn’t under a meaning-shift simply because a “possibly” was put in front of it. (In any case, this is what we are strongly inclined to say.) We know that in “possibly, Socrates did not drink hemlock”, the sentence “Socrates did not drink hemlock” doesn’t contribute its truth-value. So in that context, it doesn’t *refer* to a truth-value. Unless we assume

that “possibly” induces a meaning-shift, we must say that “Socrates did not drink hemlock”, when it occurs on its own, does not refer to a truth-value. By parity of reasoning, “snow is white”, when it occurs in a truth-functional context, does not semantically contribute a truth-value.

Of course, some *have* argued that “snow is white” (and all other sentences) undergo just that kind of meaning shift.

But whenever that strange position is argued for, it is always because it is *already* taken for granted that, in truth-functional contexts, sentences semantically contribute, or refer to, their truth-values. Frege believed, for reasons we have seen to be less than cogent, that “snow is white” refers to the truth-value *true.* So he was compelled to say that, in the sentence “possibly snow is white”[213], the sentence “snow is white” undergoes a semantic-shift.

It is true, of course, that some contexts are truth-functional and some are not. But this does not mean that sentences mean one thing in truth-functional contexts and a different thing in non-truth-functional contexts. A more innocuous line can be taken. When you say “snow is white”, you are really making a statement about the proposition *that snow is white*; you are saying that that proposition is true. Granted, the words “is true” are not in plain view. But their phonetic absence is compensated for by assertoric force. When you say “possibly, snow is white”, you are saying that that proposition is *possibly* true. When you are saying “snow is not white” you are saying that that proposition is not true. And so on. There is no need to suppose that a semantic shift occurs; what changes is not the meaning or semantic contribution of “snow is white”; what changes is what property is being ascribed to it.

Some will balk:

There is a difference, if only a subtle one, between saying “snow is white” and “*that snow is white* is true” (or, what is more colloquial, “it is true that snow is white”). When you say “snow is white” you are attributing a color to a certain kind of physical object. When you say “it is true that snow is white”, you are attributing truth to a proposition. So I must disagree with your point that when someone says “snow is white”, that person is attributing truth to a proposition.

I believe this involves a confusion.

Let us make a preparatory point. There is a difference between *saying* and *expressing.* If you say “snow is white”, you are *expressing* the fact that you believe that snow is white. But you are not *saying* that you believe that snow is white. That is why “snow is white but I don’t believe that snow is white” is not technically a self-contradiction. What the one sentence *says* does not contradict what the other sentence says. What the one sentence says contradicts what the other sentence expresses.

Consider your belief that snow is white. Here you have an attitude towards a proposition – a propositional attitude. The attitude is one of assent or belief. The proposition is: *that snow is white.* So the propositional attitude – the belief -- is to the effect that: the proposition *that snow is white* is true. In other words, *what you believe* is: the proposition *that snow is white* is true.

Beliefs are thus attitudes towards *propositions.* To believe that snow is white is to have an attitude – an attitude of assent or belief – towards the proposition *that snow is white.* Of course, having that belief involves having beliefs about things other than an entire proposition; it involves having beliefs about snow and whiteness and perhaps other things. But one has *those* beliefs by way of having a belief about the whole proposition.

Obviously we can, at least sometimes, *say* the things that we believe. And

obviously we can, at least sometimes, use sentences in a completely literal way to do that. Suppose I believe that snow is white, and I want to *say* what I believe. In that case, if I speak English, I say “snow is white”. What I believe is: the proposition *that snow is white* is true. So what I am saying is: the proposition *that snow is white* is true.

A confusion must be averted.

When I utter “snow is white”[214], I am indeed, saying something I believe. But this distinct from saying *that* I believe it. Saying what you believe is different from saying that you believe it. If I say “snow is white”, I am not *saying* that I believe that snow is white. I am *expressing* this, but not saying it. But when I say “snow is white”, I am saying what I believe (not *that* I believe it); and, as we have seen, what I believe is: the proposition *that snow is white* is true.

So, contrary to what the objector says, one *is* attributing truth to a proposition when one says “snow is white”. When you say[215] “snow is white”, you are saying something you believe. And what you believe is: the proposition *that snow is white* is true. So when you say “snow is white” you *are* attributing truth to a proposition. Of course, *in* attributing truth to that proposition, you are also attributing whiteness to snow. But you are making that attribution by way of attributing truth to a whole proposition.

Chapter 16. what are sentences?

Here is a preliminary analysis of what sentences are that, I think, has some plausibility. We will refine it later, in light of some points we will make about merits of token-semantics.

A sentence can be seen as an expression referring to a proposition that is

prefixed by a force-operator.

Let us invent a language; call it English*. English* is just like English except in a few very limited respects. In English*, there are (phonetically realized) force-operators. Let A be the assertion-operator, Q be the question-operator, and I be the imperative operator.

In English*, if you want to affirm a proposition, you take an expression that refers to that proposition, and prefix it with the assertion operator. So you don't say "snow is white". You say: "A[that snow is white]." (We can leave aside exactly how the "A" is to be pronounced.) And you don't say "is snow white?"; you say: "Q [that snow is white]". So take an expression denoting a proposition. Prefix it with the right force-operator. And that is how, in English*, you ask it, assert it, order it, and so on.

It seems to me that there is no *significant* difference between English and English*. The only difference lies in how the force-operators are phonetically realized. Obviously the English*-translation of "snow is white" is "A[snow is white]". So the natural thing to conclude is that, in English, "snow is white" consists of "that snow is white" prefixed by a force-operator. The force-operator is expressed by dropping the "that". The point is that "snow is white", and "is snow is white?" are plausibly seen as "that snow is white" prefixed by the a force-operator. In general, sentences are terms, prefixed by force-operators, that refer to propositions. We simply have to be aware that the force-operator usually doesn't consist of some discrete phonetic unit; it might consist of scrambling the word-order (as in questions) or dephoneticizing the subject (as in imperatives) or dephoneticizing the complementizer (as in assertions, questions, and imperatives). This view, it seems to me, is more natural, and more consonant with the considerations stated in this work, than the view that sentences refer to truth-values.

Chapter 17: the relation between “logical equivalence” and co-reference

I would like to clear up a confusion. Doing so will forestall some attacks on my position. It also provide us with yet another reason to reject at least some Slingshot-style arguments.

Sometimes we say that two *sentences* are logically equivalent. And sometimes we say that two *expressions* are logically equivalent.

We say “1+1=2” and “triangles have three sides” are logically equivalent.

We also say “Plato” and “the unique thing x such that x=Plato” and “the unique thing x such that 1+1=2 and x=Plato” are all logically equivalent.

In so far as we speak this way, we are guilty of an egregious equivocation. It is one thing to say that “Plato” and “the unique thing such that x=Plato” are logically equivalent. It is an entirely different thing to say that “1+1=2” and “triangles have three sides” are logically equivalent.

Slingshot-style arguments always use the premise that “logically equivalent expressions co-refer”. But this premise is quite ambiguous. For the moment, I leave it open whether this ambiguity affects the cogency of all Slingshot-style arguments. (I think it affects of *some*, but not all of them.)

Consider the expressions:

(a) “Plato”

and

(b) “the unique thing x such that 1+1=2 and x=Plato.”

(In this context, we are assuming that definite descriptions are referring

terms. If you balk at that assumption, simply replace “the” with Salmon-Kripke’s “ze”.)

When we say that (a) and (b) are “logically equivalent”, what do we mean? We *don’t* mean that the things they refer to are logically equivalent. Plato is not logically equivalent with anything. Plato is not a proposition. Nothing entails, or is entailed by, Plato.

So what do we mean when we say that (a) and (b) are “logically equivalent”. Some background would be helpful.

Let P be the condition that, given the semantic rules of English, a thing must satisfy to be picked out by (a). And let P* be the corresponding condition for (b). P entails, and is entailed by, P*. (More exactly, for all values of x, P entails and is entailed by P*.)

Given the semantic rules of English, what condition must a thing satisfy to be picked out by “Plato”. It must satisfy the condition: *O is identical with Plato* or *O is a thing x such that x is identical with Plato.* This is an open-proposition. Let P be this condition, this open-proposition.

What condition must a thing satisfy to be picked out by (b)? It must satisfy the open-proposition: *O is a unique thing x such that 1+1=2 and x is identical with Plato.* This condition is given by an open-proposition. Let P* be that open-proposition.

When we say that (a) and (b) are “logically equivalent”, we mean that P and P* are logically equivalent (or, more exactly, that for all values of x, they are logically equivalent).

Let us generalize this claim. Let us say that P is the “satisfaction-condition” for (a) and that P* is the “satisfaction-condition” for (b). Satisfaction-conditions are given by open-propositions. Every referring term has a satisfaction-condition. The satisfaction condition for “the inventor of bifocals” is given by the open-proposition: *x uniquely invented bifocals.*

[216]

The satisfaction-condition for “Socrates” is *O is identical with Socrates* or *O is a unique thing x such that x is Socrates.*

When philosophers say of two sub-sentential expressions that they are logically equivalent, what they mean is that, for all values of x, the satisfaction for the one entails and is entailed by, the satisfaction condition for the other. What is *not* meant is that the things *referred* to entail each other. (a) and (b) both refer to Plato; but Plato doesn’t entail anything, not even himself.

The kind of “logical equivalence” just described is not the only kind. (Indeed, it is not the primary kind.) So let us call it “S-logical equivalence” or simply “S-equivalence”. Two expressions are S equivalent if (for all values of x) the satisfaction condition for the one entail, and is entailed by, the satisfaction condition for the other.

Consider the expressions

(c) “that 1+1=2”

and

(d) “that triangles have three sides”.

The things that (c) and (d) *refer* to are logically equivalent. I am not saying the things (c) and (d) refer to are S equivalent – they are not. I am saying that referent of the one expression entails, and is entailed by, the referent of the other expression.

The referents of (c) and (d) are obviously not S equivalent; for S equivalence applies to expressions, not to the things meant by expressions –

not, for example, to propositions.

I would like to define a term that applies to pairs of expressions like (c) and (d). Let that term be “R-logical equivalence” or simply “R-equivalence”. (c) and (d) are R- equivalent because the things they *refer* to entail each other. In general, two expressions are R equivalent iff the things they refer to entail each other.

R-equivalence and S-equivalence are entirely different things. In fact, two expressions that are S-equivalent are *typically* not R-equivalent. (a) and (b) are obviously not R-equivalent. Neither are “the number that comes after two” and “the number that comes before four”. For the number three doesn’t entail anything.

Some S-equivalent expressions are R-equivalent: “the proposition that answers the question of how many sides a triangle has” and “the proposition that answers the question of what the square of two is”. But since most referring terms don’t refer to propositions, it follows that S-equivalence and R-equivalence tend to pull apart.

Let us focus on (c) and (d). Clearly those expressions are R-equivalent? But are they S-equivalent?

They are not. The satisfaction-condition for (c) is:

(cs) x is identical with the proposition that 1+1=2.

The satisfaction for (d) is:

(ds) x is identical with the proposition that triangles have three sides.

It is *not* the case that, for all values of x, (cs) entails, and is entailed by,

(ds). In fact, it is the case that, for all values of x, (cs) does *not* entail, and is *not* entailed by, (ds).

This is a subtle point. The proposition that 1+1=2 entails, and is entailed by, the proposition that triangles have three sides.

But (for all values of x) the proposition *x is identical with the proposition that 1+1=2* does not entail, and is not entailed by, the proposition *x is identical with the proposition that triangles have three sides.*

If (c) and (d) were to be S-equivalent, the proposition *x is identical with the proposition that 1+1=2* would have to entail, and be entailed by, the proposition *x is identical with the proposition that triangles have three sides.*

Those two propositions do not entail each other. So (c) and (d) are not S-equivalent.

(c) and (d) are R-equivalent, but not S-equivalent. (a) and (b) are S-equivalent, but not R-equivalent.

According to my analysis of sentences, the sentence

(e) “1+1=2”

refers to the same thing as (c). And, according to my analysis,

(f) “triangles have three sides”

refers to the same thing as (d).

(c) and (d) do not co-refer. So, according to my analysis, (e) and (f) do not co-refer.

This seems to conflict with the dictum that “logically equivalent expressions co-refer”. But it does not, at least not on the relevant reading of that dictum.

What is always meant by that dictum is that expressions like (a) and (b) must co-refer. In other words, what is meant is that *S*-equivalent expressions must co-refer. What is *not* meant is that (a) and (b) must *refer* to things that are logically equivalent. That would be absurd; what (a) refers to is *not* logically equivalent with what (b) refers to; for Plato isn't a proposition and therefore isn't logically equivalent with anything, even himself.

(c) and (d) are not S-equivalent, as we saw a moment ago. Any analysis that makes them be S-equivalent is wrong.

When it is said that two sentences are "logically equivalent", S-equivalence is not what is meant. When people say that "1+1=2" and "triangles have three sides" are logically equivalent, they are not saying that the satisfaction condition for the one entails, and is entailed by, the satisfaction condition for the other. They are not saying that, for all values of x, the open-proposition a thing must be meet to be referred to by the one entails the open-proposition that a thing must meet to be referred to by the other.

When it is said that, for example, "1+1=2" and "triangles have three sides" are logically equivalent, what is meant is that the proposition semantically encoded in the one entails, and is entailed by, the proposition semantically encoded in the other. In other words, what is meant is that the proposition semantically contributed by the one – the proposition that is its literal meaning – entails, and is entailed by, the proposition that is semantically contributed by the other.

The proposition semantically contributed by (c) entails, and is entailed by, the proposition semantically contributed by (d). This appears to be exactly the sense in which "1+1=2" and "triangles have three sides are "logically equivalent". (c) and (d) are *R*-equivalent. (e) and (f), it appears, are equivalent in the sense in which (c) and (d) are equivalent. In other words, (e)

and (f) are R-equivalent. R-equivalence does not guarantee co-reference; in fact, R-equivalence usually involves a divergence of reference.

We analyzed reference in terms of semantic contribution. (c) and (e) semantically contribute the same thing; so if we were right about reference, then they are refer to the same thing. Likewise for (d) and (f). (e) and (f) are R-equivalent; and they plainly and uncontroversially do not co-refer. Therefore, if our analysis of reference was right, the "logical equivalence" that holds between (c) and (d) is R-equivalence, and does not guarantee co-reference.

So a correct semantics for sentences does *not* require that logically equivalent sentences co-refer. For when we talk about "logical equivalence" in connection with sentences, we are not talking about S-equivalence; we are talking about R-equivalence. And R-equivalence has nothing to do with co-reference. Indeed, if what we said a moment ago is at all close to the truth, when people say that two sentences are "logically equivalent", they are talking about R-equivalence; they are talking about the equivalence of the things semantically contributed by – i.e. referred to by – those sentences; they are not talking about S-equivalence; they are not talking about the kind of relation that holds between (a) and (b); they are not talking about any kind of relation that requires co-reference or, indeed, that even allows for it in more than a few special cases. So when people say that two sentences are "logically equivalent", they are *not* saying, even implicitly, that those two sentences ought to co-refer. For they are talking about R-equivalence, and not S-equivalence. Indeed, what they are saying requires that those two sentences *not* co-refer, except in the special case where they both encode the very same proposition.

using this point to develop another anti-Slingshot

Many Slingshot-style arguments explicitly assume that logically equivalent *sentences* must co-refer. This principle is held to be a special case of the principle that logically equivalent *expressions* co-refer.

We've seen that this assumption involves a massive non-sequitur. It is, at best, unwarranted and, at worst, false.

The sense in which "1+1=2" and "triangles have three sides" are logically equivalent is simply *not* the sense in which "Plato" and "the unique x such that x=Plato" are logically equivalent. The first two are R-equivalent. The second two are S-equivalent.

R-equivalent expressions *do not* typically co-refer. This is uncontroversial. "That 1+1=2" *does not* co-refer with "that triangles have three sides". (Admittedly, Stalnaker holds, or used to hold, that they *do* refer to the same proposition.[217] But, for reasons lying outside the scope of this work, it is not a plausible position.)

So the kind of logical equivalence that obtains between "1+1=2" and "triangles have three sides" does *not* guarantee co-reference.

Many Slingshot-style arguments assume that that kind of logical equivalence *does* guarantee co-reference. Consider Davidson's argument (quoted from Salmon):

Let S and S* be any two sentences that have the same truth-value.

(1) S
(2) The class of all things x such that (x=x and S) is identical with the class of all things x such that (x=x).
(3) The class of all things x such that (x=x and S*) is identical with the class of all things x such that (x=x).

(4) S*.

According to Davidson, this sequence of sentence (or, more exactly, sentence-schemata) shows that S and S* must co-refer. To generate this result, Davidson needs that (1) and (2) corefer, and that (3) and (4) co-refer. He says that they do co-refer; and his grounds are that "logically equivalent" sentences co-refer.

Davidson assumes that, for any values of S and S*, (1) and (2), and (3) and (4), are logically equivalent. (As we saw, this assumption of his turned out to be wrong. But for the sake of argument, let us suppose that Davidson was right to assume this.) And on these grounds, Davidson said that (1) and (2), and (3) and (4), must co-refer.

But even if (say) (1) and (2) are logically equivalent, it does not follow that they are S-equivalent. And it is S-equivalence that is needed to guarantee co-reference. What seems overwhelmingly likely is that (1) and (2), and (3) and (4), are R-equivalent. And R-equivalence does not guarantee co-reference. Indeed, R-equivalent expressions typically *don't* co-refer.

Church's argument against Carnap

In his *Introduction to semantics,* Carnap argued that sentences refer to propositions. Alonzo Church shoed that, given principles that Carnap accepts, sentences must refer to truth-values. One of those principles is: logically equivalent sentences co-refer.

Church's argument is this. Let S and S* be any true, non-analytic sentences.

(i) S

(ii) The class of all things x such that (x=x and not S) is identical with the null class.
(iii) The class of all things x such that (x=x and not S*) is identical with the null class.
(iv) S*.

Supposedly (i) and (ii) are logically equivalent; so are (ii) and (iii); and so are (iii) and (iv). So If we grant logically equivalent sentences co-refer, then, supposedly, it follows that any two true sentences co-refer.

Church's argument is not generally taken as a mere internal criticism of Carnap; it is taken as a proof that sentences *do* refer to truth-values. This is how Church takes it in that very article.[218] He accepts Carnap's premises.

There are several reasons why Church's argument fails. Here I will focus only one.

"Plato" and the "the unique x such that 1+1=2 and x is Plato" are "logically equivalent".

But the things they semantically contribute – their referents – are not logically equivalent; Plato isn't logically equivalent with anything.

"1+1=2" and "triangles have three sides are "logically equivalent" in a different sense: the things they semantically contribute – the proposition *1+1=2* and *triangles have three sides* – are entail each other.

The second kind of logical equivalence has nothing to do with co-reference. The first kind does.

If reference reduces to semantic contribution, as we've argued it does, then the second kind of logical equivalence guarantees *lack* of co-reference, except in extremely special cases.

The Carnap-Church principle – logically equivalent sentences co-refer --

is probably false. And, contrary to what has always been assumed, that principle certainly doesn't follow from the fact that expressions like "Plato" and "the unique x such that x is Plato" co-refer. That unwarranted and probably false assumption is instrumental to Church's (1943) argument that sentences refer to truth-values. So that argument is less than cogent.

Quine's[219] slingshot – his proof that all true sentences have the same modal status -- fails for the same reason. It uses the principle that logically equivalent *sentences* co-refer.

Here is his argument. "necessarily" is obviously PSLE. Let "Phi" be any PSLE connective. And let S and S* be any two true sentences.

1. Phi (S)
2. Phi (the class of all things x such that [x=x and S] is identical with the class of all things x such that [x=x].
3. Phi (the class of all things x such that [x=x and S*] is identical with the class of all things x such that [x=x].
4. Phi (S*).

Quine thinks that (1) and (2) are logically equivalent. (He is wrong about this, as we saw. But for the sake of argument, let us assume he is right.) Given this, Quine says they must co-refer. He says the same thing about (3) and (4).

If the argument is to go through, (1) and (2) must co-refer, and so must (3) and (4).

But logically equivalent sentences don't co-refer. In any case, even if they do, that fact cannot be derived from the fact that expressions like "Plato" and "the unique x such that x=Plato" must co-refer. It would have to be established on independent grounds. Such grounds have never been provided

(and they never will be). So Quine's argument fails.

According to Gödel[220], the slingshot shows *either* that there is one fact *or* Russell's theory of descriptions is correct. Given this ugly choice, Gödel rightly infers that Russell's' theory is correct.

But there is no need to make that choice. Gödel's slingshot involves the muddle we've been discussing.

Let "Fa" and "Gb" be any two true sentences. Assume that a is not identical with b. Finally, assume that the meaning of an expression is the thing it refers to.

1. Fa. *premise*
2. a is the unique thing x such that [Fx and x=a]. *corefers with 1 by LL*
3. a is the unique thing x such that [a is not b and x=a]. *corefers with 1 by LL and CR*
4. a is not identical with b. *corefers with 3 by LL*
5. b is the unique thing y such that [y is not identical with a and y=b]. *corefers with 4 by LL*
6. b s the unique thing y such that [Gy and y=b]. *corefers with 5 by CR*
7. Gb *corefers with 6 by LL.*

If cogent, Gödel's argument shows that *either* definite descriptions must parse out, in the way described by Russell's theory, *or* that any two true sentences have the same meaning. If Russell's theory is right, then CR cannot be used to generate the needed transitions. Obviously not all sentences have the same meaning. So Russell's theory must be right.

There are may problems with Gödel's argument. Here let us focus on one. Obviously that argument relies heavily on the idea that logically equivalent sentences co-refer. This assumption is unwarranted – it appears warranted

only when S-equivalent and L-equivalence are conflated -- and it is probably false. So Gödel's argument fails.

Summary

The kind of "logical equivalence" that guarantees co-reference is S-equivalence, not R-equivalence. When it is said that two sentences are logically equivalent, what is meant is *not* that they are S-equivalent. So a correct semantics doesn't necessarily make sentences like "1+1=2" and "triangles have three sides" co-refer. So it is not to the discredit of our analysis that it makes those two sentences refer to different things.

Indeed, that fact is to the credit of our analysis. The relation that holds between those two sentences seems to be R-equivalence. And R-equivalence does *not* involve co-reference. Indeed, most pairs of R-equivalent expressions do not co-corefer. So, if anything, a correct semantics makes most pairs of logically equivalent sentences *not* co-refer; it makes "1+1=2" and "triangles have three sides" refer to different things. A correct semantics makes any two sentences that bear different propositions *not* co-refer, regardless of whether they entail each other or not. Our analysis satisfies that desideratum. [221]

Chapter 18

Carnap (1947) said that that the intension of a "sentence" (he didn't distinguish tokens from types) is a proposition, and that extension of a sentence is a truth-value.

This view is generally seen as either true or -- if not quite true – then very reasonable: an admissible extension (no pun intended) of the term "extension".

Carnap's view, I am afraid, doesn't quite deserve such a warm verdict.

Let us start with a more straightforward case of a word that is alleged to have both extension and intension.

The word "whale", we are told, has both intension and extension. The intension is a property or function (from objects to the truth-values). The extension is the class of whales (Shamu, Willy, Moby Dick…)

But there is no one expression that has both intension and extension. Tokens of "whale" have extension; the expression-type "whale" has intension. Nothing has both.

We don't want to say that the expression-type "whales" denotes *different* classes in different worlds, while having the same semantics. That is perilously close to absurdity. The *extension* of a term is, in effect, the set of objects that it refers to. So unless we divorce reference from semantic content, it is self-contradictory to say that "whale" has the same semantics in worlds where it has different extensions.

What we must say is: *tokens* of "whales" denote different classes in different worlds; the *type* has an invariant meaning: it pairs off a token of "whale" in a world W with a class C iff C is the class containing all and only whales in W.

The intension of "whale" is a function from worlds to classes. Tokens of it have *extensions.* In this world (at this time), it has as its extension the class containing all and only: *Shame, Willy.…*

By parity of reasoning, sentence-types have *intension*; sentence-tokens have extension.

The intension of a sentence is some kind of function. But it cannot be a function from worlds to truth-values, for reasons we've seen. The type "I am tired' does not bear any proposition; so it cannot have a proposition for its intension. The same must be true of the type "Socrates was wise". If assign a

different kid of intension to the type “Socrates was wise”, we end up committed to the highly arbitrary view that sentences can have entirely different sorts of things for the intensions – some have propositions (functions from worlds to truth-values), others have functions from contexts of utterance to truth-values.

In so far as the type “I am tired” has a meaning, it is a function from contexts of utterance to propositions. So if anything is its intension, it is that function. If anything is the intension of “Socrates was wise”, it is a function from contexts (worlds and possibly times) of utterance to propositions.

The *extension* of a token of “I am tired” must be the thing that the *intension* of the corresponding type assigns to that token – just as the *extension* of a token of “whale” is the thing that the intension of the corresponding type assigns to that token. So the extension of “I am tired” is a proposition. Thus, the extension of “Socrates was wise” is a proposition – not, *pace* Carnap, an extension.

Appendix A. *an argument of Davidson’s*

Everybody thinks it *strange* to suppose that “snow is white” means one thing on its own and another in “John believes that snow is white”. But is usually held that this supposition is coherent.

Donald Davidson (“Theories of Meaning and Learnable Languages”[222]) put forth a clever argument that challenges this.

I do not know whether Davidson’s argument is decisive. But if cogent, it is absolutely devastating. Yet I’ve never seen so much as mentioned in passing in Slingshot-literature. (Neale[223] never mentions it.) In any case, it

deserves to be mentioned.

Davidson's argument is this. English is obviously a learnable language. Any learnable language has a finitely large primitive lexicon. If the Frege-Church line is correct, then "snow is white" means one thing on its own, a second thing in "Tim said that John believes that snow is white". And so on.

If we abstract from fatigue, limitations of memory, and the like, it is pretty clear that, in virtue of your knowledge of English semantics, you can understand *all* of these sentences. No matter how deeply embedded "snow is white" is, you will be able to understand it. It might be pre-fixed by 42,345 epistemic operators; you will still be able to understand it.

Granted, it is probably not humanly possible to understand such a sentence. But that has nothing to do with your not being able to assign meaning to "snow is white" in that sentence; it is not a *semantic* limitation on your part; it is a limitation -- set by things like fatigue, memory-limitations, and the like to *exploit* -- your semantic knowledge.(It is a performance-problem, not a competence-problem.[224])

Let "snow is white$_1$" be synonymous with the sentence "snow is white" when it occurs on its own, or in an otherwise truth-functional context. Let "snow is white$_2$" be synonymous with "snow is white" when the latter occurs in the scope of a single epistemic operator. Let "snow is white$_3$" be synonymous with "snow is white" when the latter occurs in the scope of two epistemic operators. And so on.

Anyone can understand "snow is white$_{7,865}$".If Frege is right, it means something totally different from "snow is white$_1$" and "snow is white$_{978,787}$". If Frege is right, for any i, your knowledge of English semantics comprises a knowledge one semantic rule for "snow is white$_i$", and a different one for "snow is white$_{i+1}$". So your knowledge of English

semantics comprises *infinitely* many semantic rules, corresponding to “snow is white”.

By itself, this might seem acceptable. After all, English semantics *does* comprise infinitely many rules: but most of them are *derived*; and one can learn those on the basis of the primitive rules.

But this kind of consideration does not apply here, as Davidson points out. Frege’s theory, by implication, makes “snow” be *infinitely* ambiguous. There is “snow_1”, “snow_2”, “snow_3”, and so on. The physical-type “snow” is ambiguous between all these different words. To *really* learn the meaning(s) of “snow”, one must learn the meanings of each of these disambiguations. Otherwise, one won’t know what “snow” means in a sentence like “Wayne believes that Terry hopes that Michael dreads that Oliver hopes that Lenny desires that Sarah opines that John believes that snow is white”. But one always *does* know what “snow” means in such contexts, even if one is otherwise unable to parse that sentence.

So Frege is saying, at least by implication, that each of “snow_1”, “snow_2”, “snow_3”… *does* have a different meaning. If Frege is right, each English speaker *does* understand each of these infinitely many different words (or, depending on how you define “word”, each of these infinitely many disambiguations of a single word).

For each i, “snow_i” is a *primitive* entry in our lexicon. It has no constituent structure; no recursive rules give it meaning.

Thus Frege’s analysis requires that each person who speaks a natural language know *infinitely* many primitive semantic rules. This is not acceptable; for it makes natural languages be unlearnable.

Admittedly, this argument is not airtight. One might say:

Maybe each of “snow_1”, “snow_2”, “snow_3“…are primitives. But that

doesn't mean that we must simply *memorize* all of their meanings. Perhaps it means that we cannot derive their meanings in quite the way we derive the meaning of a complex expression (say "the big ugly man didn't say hello to me"). But there are other ways of figuring out what an expression means.

Obviously we can learn the meaning of "snow" (i.e. "snow_1"). Let us suppose, as Frege did, that this involves our knowing some sense that applies to snow. (This sense is the meaning of "snow_2".) Since, by hypothesis, we grasp some sense S that applies to snow, perhaps we can *figure out* what sense S* that applies to S; and then we can figure out some sense S** applies to S*; and so on. We would figure this out not via semantic rules, but via general epistemological or logical considerations. So even though, each of "snow_1", "snow_2"…are semantically primitive, we can still figure out what they mean.

But this runs afoul of Russell's plausible point that there is no "backwards road" denotation to sense. Any given thing can be given by *infinitely* many senses. This means that one of two things happens to Frege's theory.

(i) A sentence like "Fred believes that Phil thinks that snow is white" is ambiguous as many times over as there are senses that apply to snow, and senses as there are that apply to the senses that apply to snow.

But surely "Fred believes that Phil thinks that snow is white" is not ambiguous, at least not to that degree. This brings us to Frege's second option:

(ii) Frege has to choose some *specific* sense, out of the infinitely many

possible ones, that is meant by “snow”; and he has to do this for *each* of the various disambiguations of “snow” (i.e. he must do it for “snow1”, “snow2”, and so on.) But it is hard to see any such choice could fail to be arbitrary.

By choosing some *specific* sense for each of “snow_1”, “snow_2”…Frege avoids rendering that sentence ambiguous. But it is hard to see any such choice could fail to be arbitrary.

I conclude that Davidson has found a real defect in Frege’s theory; and that attempts to patch it up are unpromising.

Appendix B. the context principle, and some problems with truth-conditional semantics.

In this section, I’d like to give some historical antecedents to my analysis of reference.

Two points that are not merely of historical interest will fall out of this discussion. The first is that the main connective in semantic rules ought to be a “means that”, not a “is true iff”.

The second is that the age-old distinction between “lexical” and “non-lexical” items – between “categorematic” and “syncategorematic” expressions – is either spurious or is a distinction between different types of referring terms: “lexical” items being those that refer to individuals or first-level functions; “non-lexical” items being those that refer to second-level functions.

Dummett and the context principle

Frege famously said that it is only in the context of a sentence that a word has meaning. His exact words are:

"[N]ever as for the meaning of a word in isolation, but only in the context of a proposition".[225]

I'd like to briefly discuss this "gnomic utterance"[226]. I will briefly discuss what I think it means. Then I will briefly discuss what Dummett and Tughendat have to say. My own analysis differs little from Dummett's.

Frege wrote that in 1884 (or thereabouts). Let's jump forward to 1905. Frege's statement can, I think, be understood in terms of a famous semantic thesis of Russell's, put forth at that later date. Russell famously argued that "the phi" is not a referring expression – it doesn't denote the unique thing having phi. Instead, Russell argues, "the phi" is to be defined "contextually". You don't give the meaning of "the phi" by pointing to some object and saying: there! That individual is the meaning of "the phi". You give the meanings of *whole sentences* containing that expression. "The phi" doesn't have "any meaning in isolation". But whole sentences containing it have meaning. And you give the semantics of "the phi" by giving a rule assigning meaning to such sentences. "The phi" is thus to be defined "contextually". (The rule in question is roughly: "…the phi…" means: exactly one thing x has phi and, for any y that has phi,…y…)

This thesis of Russell's was really just an installment in a program begun by Frege. Frege showed that "exactly one man" is not a referring expression. It is to be defined contextually.

"…exactly one man…"

means:

for some x, x is a man and…x…and for any y such that y is a man and…y…, y=x.

So "exactly one man snores" means: for some x, x is a man and x snores and for any y, if y is a man and y snores, then y=x.

Thus "exactly one man" is to be defined *contextually*. It is defined by giving the meanings or truth-conditions of whole sentences in which it occurs. It is not defined by associating it with some object; it is not defined in the way that, supposedly, proper names are defined.

Frege argued compellingly that many quantifiers – "all", "some", "some man", "no man" – are to be defined contextually. Russell was trying to show that what Frege said about "some phi", "all phi's", and so on, was true (mutatis mutandis) of "the phi".

Frege's famous utterance -- "never ask after the meaning of an expression outside the context of a sentence" – occurs at the beginning of *The Foundations of Arithmetic*. One of the two or three major points of that book is that grammatical form and logical form may diverge.[227] And, in that book, the divergence is always illustrated in connection with *quantified* sentences. Grammatically, "exactly one man" has the same role as "Smith". But there is a difference. "Smith" is defined by associating it with some individual. "Exactly one man" is defined contextually, by showing what is meant by whole sentences containing it.

This, it very much seems to me, is at least part of what Frege had in mind

when he put forth his famous dictum. He was alluding to his insight that logical and form and grammatical form pull apart quite dramatically when it comes to *quantified* sentences.

Incidentally, Wittgenstein argued that "and" and "or" and "not" are not referring terms and that *they*, like quantifiers, are to be defined contextually. The meaning of "and" is not given by indicating some (platonic) object; it is given by some rule like "S and S*" iff S is true and S is true. (We have seen that "and" *is* plausibly seen as denoting a platonic object, contrary to what Wittgenstein thought.)

Frege's point that "some man", "two men", and so on, do not refer to individuals is, of course, correct. And so is his point that such expressions are to be defined contextually. "…some man…" means: for some x, x is a man and…x… [228] So given what Frege said, it seems that the meanings of at least some expression are to be found in the contributions they make to sentences in which they occur; it would be nonsense to assign them meanings "in isolation" of sentential contexts.

What is not often realized is that, at least arguably, *all* expressions -- referring expressions no less than quantifiers -- must be defined contextually.

Arguably, "John" is no more capable of a *non*-contextual definition than "exactly one man". Of course, it "John" certainly *seems* to be capable of a non-contextual definition. We can define "John" by *pointing* to him; we don't need to say what is meant by whole sentences of the form "…John…"

But, I would argue, when we define "John" by pointing to him – when we say "*that* man over there is John" – we are really saying:

sentences of the form "…John…" mean:…*O*…,

where O is the man indicated by the "that".

It is this point that lies at the basis of *Dummett's* analysis of Frege's famous dictum.

Dummett's interpretation of Frege is like mine with one qualification: the point Dummett attributes to Frege is a generalization of the point I just attributed to Frege.

According to Dummett, what Frege means is this. An expression "E" means M iff M is what E contributes to the truth-conditions of sentences in which it occurs. Dummett gives a two-step argument for this. Here is the first step:

> To assign a reference to a name or a set of names…could only have significance as a preparation for their use in sentences; analogously the assignment of a reference to physical objects could only be understood as a preparation for the use of those objects within some representative arrangement. If, for example, I take some coloured counters, and say, 'Let this one stand for th Government, this one of the Opposition, this one for the Church, this one for the Universities, this one for the Army, this one for the trade unions,…", and so on, I shall be understood on the presumption that I am about to make some arrangement of the counters by means of which I intend to represent some relations between these institutions, and assert that they obtain. If I do not go on to make any such arrangement, but simply start talking about something else, my earlier declarations lose their original intelligibility; I cannot, when questioned why I said all that, replay, "oh, I just wanted to hose counters to stand for those things, that's all'; for their standing for those thing sonly amounts to anything if they are to be used to effect some symbolic representation by means of which a thought is expressed. Otherwise, my stipulation of their reference is like my saying, in the course of explaining a card game, 'Ace is high', and it later turns out that

the ranking of the cards plays no role in the game; or it is like my saying, 'Suppose there is life on Mars', and then failing to draw any consequences from this hypothesis, and, when challenged, saying, 'Oh, I simply wanted you to suppose'; or, perhaps, it is most like giving a definition of a word and then not subsequently using that word…More generally, the assignment of a sense to a word, whether a name or an expression of any other logical type, only has significance in relation to the subsequent occurrence of that word in sentences. A sentence is, as we have said, the smallest unit of language with which a linguistic act can be accomplished, with which a 'move can be made in the language-game': so you cannot *do* anything with a word – cannot effect any conventional (linguistic) act by uttering it- save by uttering some sentence containing that word (save for cases [in which a single word is elliptical for a whole sentence]). (Dummett 1973: 193-194)

The purpose of language is to say things. Nothing linguistic has been accomplished until something has been said. Sub-sentential expressions don't say anything. Only whole sentences do that. Of course, sub--sentential expression are bits of language. But that is only because they are *constituents* of things that say anything. It is only because they are constituents of sentences.

Given this, it is a very short step to saying: the meaning of a sub-sentential expression *is* the contribution that it makes to the meanings of whole sentences of which it is a constituent.

One might balk at what we've just said:

It might seem that if I say "Socrates is a", and then stop, I *have* done something linguistic. Granted, I've produced something ill-formed. But it is an ill-formed bit of *language.* This destroys your point.

This doesn't really affect the point. An ill-formed utterance is not a linguistic *accomplishment*. It is an attempt at such an accomplishment. But it is no more such an accomplishment than *falling* to run a marathon is actually running a marathon. Only whole sentences constitute linguistic *accomplishments.* So the smallest unit of linguistic accomplishment is the whole sentence. There are no sub-sentential linguistic accomplishments. There is sub-sentential linguistic meaning. But for an expression to have such meaning to have a certain role in linguistic-sentential accomplishment. For a sub-sentential expression to have meaning is for it to contribute to the meanings of sentences.

Suppose I say "let 'A' be the name of the number one, let 'B' be the name of the number two…" The understanding is that we are going to use "A" and "B" to make *statements* about those numbers. Suppose I say "I *just* want 'A' to stand for one, and 'B' to stand for two…But those symbols are not to be used in the construction of statements *about* those numbers." What I've said is incoherent. There is *no* sense in which "A" can stand for the number one unless "A" can be a part of sentences that make statements about that number.

Once that is granted, we are committed to saying: for "A" to stand for the number one just *is* for "A" to make a certain kind of contribution to sentences which make statements about that number.

Given this, the salient question is: What is the nature of that contribution? Dummett answers:

[T]he sense of a word or expression always consists in the contribution it makes to determining the thought [the proposition] expressed by a sentence

in which it occurs. That is: the sense of the word consist in a rule which, taken together with the rules constitutive of the senses of other words, determines the condition for the truth of a sentence in which the word occurs. The sense of a word thus consists – wholly consists – in something which as a certain relation to the truth-value of sentences containing that word." (Dummett 1973: 193-194).

Let us make this point a little more concrete. For the sake of argument, suppose that the meaning of "Plato" is simply Plato himself.[229] According to Dummett, what this supposition amount to is: when "Plato" occurs in a sentence, Plato himself is the contribution made to the truth-conditions of that sentence.

But what does *that* mean? What exactly does it mean to say that, when "Plato" occurs in a sentence, it "contributes" Plato to the "truth-conditions" of that sentence? It is to say that sentences of the form "…Plato…" are true just in case:…*Plato*…(i.e. just in case Plato has:…x…)

"Plato taught Socrates" is true iff *Plato* had the property denoted by *x taught Socrates*. "Xenon kicked Plato" is true iff Plato had the property denoted by *Xenon kicked x*. If "Xenon kicked Plato" were true iff Xenon kicked Socrates, then (holding constant all the other semantic rules of English), "Plato" would refer to Socrates.

I do not know whether Dummett's point is correct as an interpretation of Frege. So let us refer to it not as Frege's point, but as Dummett-Frege's point.

The point I attributed to Frege is a special case of Dummett-Frege's point. On my reading, Frege is saying: *quantifiers* are to be defined contextually. Dummett-Frege is saying: not only quantifiers, but *all* expressions, are to be defined contextually.

With one qualification, Dummett-Frege's analysis of reference coincides

with *our* analysis of reference. Our analysis is: For "Plato" to refer to Plato is for "Plato" to contribute Plato to the meanings of sentences in which it occurs; more exactly, it is for "…Plato…" to mean:…*Plato*…(or *Plato has*… *x*…).

Here is where I *disagree* with Frege-Dummett's analysis. Dummett-Frege speaks of *truth*-conditions as opposed to meanings. And this makes his analysis fail.[230]

Consider the statement:

(*) any sentence of the form "…Plato…" is true iff :…*Plato*…(or *Plato has*… *x*…).

According to Dummett-Frege, (*) gives the meaning of the statement that Plato is the semantic value – the meaning – of Plato.

There is a problem with (*).What is the meaning of the "iff" in (*)? Obviously the "iff" is not the "iff" of material equivalence. For *"Socrates smoked" is true iff grass is green* is true if the "iff" is taken I that way. But plainly *grass is green* is not the semantic content of "Socrates smoked".[231]

Presumably it is the "iff" of logical equivalence. But in that case (*) gives the wrong result. The proposition on the right side of the "iff" need not be the one semantically encoded in the sentence on the left. "Plato smoked" is true iff Plato smoked. But "Plato smoked" is also true iff Plato smoked and 1+1=2. But "Plato smoked" does not mean: *Plato smoked and 1+1=2*. "Plato smoked" means simply: *Plato smoked*.[232] The truth-conditional approach has the absurd result that:

(***) "Plato smoked" means *Plato smoked and 1+1=2*.

Truth-conditions are not fine-grained enough to constitute meanings. What are the truth-conditions for “Plato smoked”? In other words, for what condition C is it the case that C’s satisfaction is necessary and sufficient for the truth of “Plato smoked” and no other sentence? The answer is: none. Any condition – any model, any world, any situation --- that suffices for the truth of “Plato smoked” will also suffice for the truth of “Plato smoked and arithmetic is incomplete”,
“Plato smoked and 1+1=2”, and so on. So (*) says, in effect, that

(***) “Plato smoked” means *Plato smoked and 1+1=2.*

Any condition sufficient for the truth of “Plato smoked” is sufficient for the truth of infinitely many other sentences with palpably different meanings.

Some propositions are true under *all* conditions. *No* conditions (or at least no particular condition) must satisfied to make them true. They will be true regardless of whether Plato smoked or not. These propositions make it impossible to give the meaning of “Plato smoked” in terms of truth-conditions. In general, these propositions will make it impossible to give a purely truth-conditional semantics for *any* sentence of the form “…Plato…” Consequently the semantics of “Plato” cannot be given by giving a rule for the *truth*-conditions for sentences of the form “…Plato….”

So, as it stands, Dummett-Frege’s point must be rejected. The right analysis, I would suggest, dispenses with truth-conditions and talks simply of meaning. For “Plato” to have Plato for its semantics is for sentences of the form “…Plato…” to *mean*:…*Plato*…. Meanings are sufficiently fine-grained that semantic rules, thus defined, rule out absurd results like (***).

There is another reason why the “iff” in semantic rules cannot denote

logical equivalence. The inscription “snow is white” could mean anything; it obviously isn’t a matter of *logic* that it means *snow is white.* It is an empirical and contingent fact.

So

(S) “snow is white” is true iff snow is white

is simply false, if the “iff” denotes logical equivalence.

Of course, a sentence is not plausibly seen as being simply a certain physical type; it is a physical type *paired off* with a certain meaning. In any case, on *one* interpretation of the word “sentence”, sentences are not mere physical types, but are pairings of such types with meanings. It might seem that, on this interpretation of “sentence”, the following *is* a logical truth:

(S1) The sentence “snow is white” is true iff snow is white.

But there are several problems. First of all, if we define “sentence” as a pairing of meanings with physical types, then the truth-conditional approach – which aspires to do away with meanings, and replace them with truth-conditions – is an immediate failure. For, if we define “sentence” in the way just described, then we are countenancing meanings. And in that case, instead of saying (S1), we might as well just say:

($S1_M$) “snow is white” *means* snow is white.

One might try to rescue the truth-conditional approach by defining a “sentence” as a pairing of *truth*-conditions with physical types. On that interpretation of the word “sentence”, the sentence “snow is white” is a

pairing off a certain physical type with a certain set of truth-conditions. But in *that* case

(S1) The sentence “snow is white” is true iff snow is white.

becomes trivial. We’ve *defined* the sentence “snow is white” as something with certain truth-conditions. But in (S1) we then go one to ascribe those very same truth-conditions to that sentence. So, on that interpretation of “sentence”, (S1) becomes a pleonasm; it becomes something which is no more informative than “unmarried males are males”. This is not acceptable since semantic rules are typically informative: an Albanian who learned the semantic rule giving meaning to “snow is white” (in English) would be learning something contingent and significant.

We’ve exposed a deep incoherence in the truth-conditional approach. Semantic rules are supposed to be *informative.* They are supposed to tell you what certain inscriptions and strings of sound mean; they are supposed to give *empirical* information. This information is not *a priori* or analytic. But if we make the main connective in semantic rules be the logical “iff”, then we are, in effect, saying that such information *is* analytic. But this makes it impossible for semantic rules to be empirically informative. It is very hard to identify *any* meaning of “iff” whereby (S1) is both true and informative.

There is another reason why (S1) is false, on any reading of “iff”. First of all, “snow is white” has different meanings. We could create a language L in which it means *grass is green.* In Albanian, it doesn’t mean anything. Meaningfulness is always relative to a language. So (S1) should really be:

(S2) The sentence “snow is white” is true *in English* iff snow is white.

But here is a major problem. A language just *is* a list of semantic rules. English just *is* a list of semantic rules. On pain of vicious circularity, those rules cannot *themselves* refer to the English. So to avoid this vicious circularity, we must drop the "in English" from (S2). But then we run into the problems we ran into earlier.

It is becoming harder and harder to escape the following conclusion: No sentence of the form "S is true (in L) iff P" can possibly claim to constitute a *semantic* rule. (S1), and variants like (S2), simply don't give the right meanings. They aren't fine-grained enough. And they turn empirical truths into (false) analytic statements. Finally, they are guilty of vicious circularity – of referring to the very language that they are supposed to constitute.

The only conceivable way to rescue truth-conditional semantics is to come up with a new kind of "iff" -- one that allowed all the desiderata we've been discussing to be satisfied; one that allowed semantic rules to be, simultaneously, empirical informative and logically compulsory; one that allowed "snow is white" to be true "iff" snow is white, but *not* true "iff" snow is white and 1+1=2. But, clearly, such an "iff" is not to be had.

The right view, then, seems to be that semantic rules ineliminably make use of the concept of *meaning.* In English, "snow is white" means snow is white. That is a semantic rule (albeit a derived one) of English. There is no way to replace the "means that" in that rule with an "iff". The "iff" will inevitably be too permissive. Indeed, the "iff" will create falsity where previously there was truth; it obviously is *not* necessary, in any relevant sense of "iff", that "snow is white" is true iff snow is white.

To sum up, with one qualification, Dummett's conception of reference is exceedingly plausible. That conception is:

"Plato" refers to Plato iff sentences of the form "…Plato…" are dependent

for their truth or falsity on Plato's having…x…

In other words,

For "Plato" to refer to Plato is for sentences of the form "…Plato…" to be true iff Plato has…x…

The only problem is that the "iff" should be replaced with a "means that".

a note on Quine on definition

A passage from Quine helps put some of these issues into logical and historical context. Quine wrote:

Definition used to be thought of primarily as equating a term outright to some expression. This view derived from the old doctrine of real definition: a term would be a name of some object or kind, and the definition would specify the essence. Some words could not be viewed as naming objects or kinds; they were called *syncategoremata,* and might be accounted for by including them in term phrases that were defined as wholes. It may have been Jeremy Bentham who first appreciated that terms generally might be handled like the syncategoremata. All we need require of a definition, for full eliminability, is that it provide for elimination of the defined word by paraphrasing all contexts in which the word the word is to be used. Such was his theory of fictions, or of paraphrases, known later as the method of contextual definition.[233]

There is a useful mixture of truth and falsity in this passage. Let us start by spelling it out.

The common-sense conception of reference is this. There is an object O; you attach a label L to O; L refers to O. The label is a word, not (say) a ribbon; and the label is attached verbally, not physically. But with that qualification, referring terms are simply labels.

The label-conception of word-meaning very much seems *not* to apply to certain words. Surely "or" and "not" and "because" don't refer to anything. Such words were called "syncategorematic" terms by the medievals; today, linguists refer to them as "non-lexical items". (Lexical items are those that have isolable meanings – "John", "red" and whole sentences. Non-lexical items are those that have a structural, not a denotative or expressive, role; they are pieces of connective material that can be used to build new lexical items out of old ones. But they themselves don't express or pick anything out. "Or" doesn't express or pick anything out; but it can be used to build new lexical items (E.g. "grass is green or snow is white") out of old ones.

The obvious thing to say about syncategoremata is that they must be defined *contextually.* You don't give the meaning of "or" by pointing to some object and saying "that is what 'or' stands for". You don't define it the way you define "Smith". You define it contextually, by its introduction or elimination rules perhaps.

Quine comes in at this point and say: *all* terms, not just syncategorematic words, can be defined contextually. One way to define "Smith" is ostensively: "*that* guy is named 'Smith'". Another way, says Quine, is contextually: supposing that O is *that* guy over there, then sentences of the form "…Smith…" mean:…*O*…

Bentham used applied the notion of contextual definition to terms like

"Hamlet" and "Zeus". We don't want to say that they pick anything out. But we do want to say that they are meaningful. We do this by showing how they parse out of sentences containing them; we define them contextually. (I myself do not know exactly *how* Bentham proposed that "Hamlet" parses out; I am going off a few passing comments made by secondary sources.)

I agree with everything Quine says, apart from one nuance. Quine seems to be saying that there is some *opposition* between contextual definition and the other kind, the "labeling" kind; they are actually *different* ways of defining an expression. In this work, I have been arguing, in effect, that there is *no* such opposition. When you point to some object O, and say "that is Smith", or "that guy's name is 'Smith'", you are, I believe, saying: "… Smith…" means:…*O*…

Of course, no three year old, and most adults, won't say "I just told you who 'Smith' names, and my doing that involved my showing you what is meant by whole sentences of the form '…Smith…'." But proves nothing. The semantic import of our everyday linguistic practices are often hidden from us; that is why semantics is so hard. That is why people thought that "some man" referred to an ambiguous man before (and even after) Frege came along.

The other side of the coin is that it is not easy to draw a line between categorematic and syncategorematic – lexical and non-lexical – items. The distinction is not to be drawn in terms of how these terms are defined. For they're all defined the same way – contextually.

Presumably, the difference is this. "Smith" picks something out, whereas "or" does not. But when you say that "Smith" picks something out, you are just saying that a sentence of the form "…Smith…" has a certain meaning. So the difference between "or" and "Smith" is not captured by saying: "Smith" picks something out, where "or" does not. For to say that "Smith"

picks something out is just a condensed way of that “Smith” admits of a certain kind of contextual definition. And “or” equally admits of contextual definition.

There is still hope for somebody who wishes to distinguish lexical from non-lexical items. They can say:

I grant that for “Smith” to refer to Smith is just for “Smith” to admit of a certain kind of contextual definition. And I also admit that “or” is to be defined contextually as well.

But there is a difference. The contextual definition for “Smith” is to the effect that the truth of certain sentences (those of the form “…Smith..”) turn on facts about a certain *individual*. Sentences of the form “…Smith…” say that *Smith* has…x…So the truth of such sentences turns on facts about some individual. The story with “or” is different. There is no individual O such that the truth of “…or…” turns on whether O has…x…

This is a good counter-move. But it still fails. When linguists talk about “lexical” items, they are not referring only to terms that pick out *individuals*; they are referring terms like “red” and “smart”. These things refer to properties or functions. Linguists characterize them as “lexical” because, intuitively, they do seem to have a denotative or expressive function. And linguists characterize “or” and “and” as non-lexical because, intuitively, they seem to *not* have a denotative or expressive function; they seem to have a syntactical or formal function.

Here is the problem. If you count “red” and “smart” as lexical items, then you must also count “or” and “and”. The first two almost certainly pick out functions or properties. We’ve already seen how. (“And” and “or” are functions from sequences of sentences to truth-values, or they are

properties of such sequences.) So “and” is no less denotative than “red”.

Granted, there is still *a* categorial difference between “and” and “smart”, even though they both pick out functions or properties. “Smart” picks out a first level function or property (a function from individual to truth-values, a property of individuals). “And” picks out a second-level function (a function from sequences of sentences to truth-values, a property of sequences of sentences).

It seems likely that morphemes like tense-markers and case-markers can be made out to denote functions or properties.

The upshot is this. *Everything* is a referring term. *If* there is a distinction between “lexical” and “non-lexical” items, it is not a difference between denotative terms and non-denotative terms; it is a difference between terms that denote certain things (individuals or first-level functions) and terms that denote other things (second-level functions). The Augustinian conception of language, ridiculed by Wittgenstein, turned out to be quite right.

Tughendat on reference

Ernst Tughendat put forth an interesting analysis of reference. He attributes this analysis to Frege.[234]

Why is it important to know the meanings of words? Because such knowledge helps us figure out what Tughendat calls the “significances” of sentences.[235]

Suppose I could know the meaning of any English sentence *without* knowing the meanings of any English words or, indeed, of any sub-sentential expressions of English. (I myself do not believe this is a coherent

supposition. But it does not have to be, given where I am going with it.) In that case, it really wouldn't matter that I didn't know the meanings of any English words. Under that circumstance, the sole which thing for knowledge of word-meaning is important – the understanding of whole sentences – would be independently available to me.

Under that circumstance, certain niceties about English semantics would be unknown to me. If I heard the sound "snow is white", I wouldn't know what meant what; I wouldn't know whether "white" meant *snow*, for example. And this would make it impossible for me to know many things *about* English semantics; for example, I couldn't do derivation trees. But to no degree would I be barred from *using* English sentences – from using them to say the things I want to say and to understand the things I want to understand. My competence *within* English would be quite unscathed. (We might say that, while my knowledge of English was impeccable, my knowledge of meta-English was limited.)

Knowledge of word-meaning is useless except in so far as it relates to knowledge of sentence-meaning. Therefore word meaning *itself* is useless except in so far as it relates to sentence-meaning. [236]

This suggests that we should try to understand the concept of word-meaning (or, more generally, of sub-sentential meaning) in terms of the concept of sentence-meaning.

This what Tughendat does. He says: for two sub-sentential expressions to have the same meaning simply is for them to make the same contribution to the meanings of the sentences in which they occur. Once this is granted, it is a short-step to saying that a sub-sentential expression E to mean O just *is* for E to contribute O to the meanings of sentences in which E occurs. Given this, it is a short step to saying: for E to *refer* to O just is for E to contribute to the meanings of sentences in which E occurs.

Bibliography

Meaning and Truth. 1991. Jay L. Garfield and Murray Kitely (eds.)
St. Paul, Minnesota: Paragon House

The Philosophy of Language. (1990.) Second Edition A.P. Martinich (ed.).
Oxford: Oxford University Press.

Barwise, Jon and Perry, John.1990/1983. "Semantic Innocence and Uncompromising Situations". In *The Philosophy of Language.* (1990.)
Second Edition A.P. Martinich (ed.).
Oxford: Oxford University Press.
-- *Situations and Attitudes.* 1999. Palo Alto: CSLI Publications.

Blackburn, Simon. 1984. *Spreading the Word.* Oxford: Clarendon.

Carnap, Rudolph. 1947. *Meaning and Necessity.* Chicago: University of Chicago Press.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge: M.I.T. Press.

Church, Alonzo. "Carnap's Introduction to Semantics". *The Philosophical*

Review. Vol. LII. No. 3. 1943. 298-304.
--1956. *Introduction to Mathematical Logic.* Princeton: Princeton University Press.

Davidson, Donald. “Theories of Meaning and Learnable Languages”. 1984/1965. In *Essays on Truth and Interpretation.* Oxford: Oxford University Press.
--1990/1967. “Truth and Meaning”. In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.

Donnellan, Keith. 1990/1966. “Reference and Definite Descriptions.” In A.P. Martinich (ed). *The Philosophy of Language*. Oxford: Oxford University Press.

Dummett, Michael 1981. *Frege: Philosophy of Language.* Second Edition Cambridge: Harvard University Press.

Evans, Gareth. 1990/1974. “The Causal Theory of Names”. In A.P. Martinich (ed). *The Philosophy of Language*. Oxford: Oxford University Press.

Frege, Gottlob. 1997/1879. “Begriffsschrift”. In *The Frege Reader*. Edited by Michael Beaney. Oxford: Blackwell.
1997/1891. “On Function and Concept.” In *The Frege Reader*. Edited by Michael Beaney. Oxford: Blackwell.
--1990/1982a. “On Sense and Nominatum.” In A.P. Martinich (ed) *The Philosophy of Language* Oxford: Oxford University Press.
-- 1953/1884. *The Foundations of Arithmetic.* Translated by J.L. Austin.

Oxford: Basil Blackwell. Reprinted in 1980 by Northwestern University Press.
-- 1997/1892b. “On Concept and Object.” In *The Frege Reader.* Edited by Michael Beaney. Oxford: Blackwell.

Gödel, Kurt. 1944. “Russell’s Mathematical Logic.” In *The Philosophy of Bertrand Russell.* Edited by Paul Arthur Schilpp. Evanston, Illinois: Northwestern University Press.
Kaplan, David (1989.) ‘Demonstratives’ in *Themes from Kaplan.* Oxford: Oxford University Press

Kripke, Saul. 1999/1971. “Identity and Necessity”. In *Metaphysics: An Anthology.* Edited by Jaegwon Kim and Ernest Sosa. Oxford: Blackwell.
1991/1977 ‘Speakers Reference and Semantic Reference’. In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.
--1972. *Naming and Necessity.* Cambridge: Harvard University Press.

Mill, John Stuart. 1991. “Of Names and Propositions”. In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.

Montague, Richard. 1974. “The Proper Treatment of Quantification in Ordinary English”. In *Formal Philosophy.* Edited by Richard Thomason. New Haven: Yale University Press.

Neale, Stephen. 1990. *Descriptions.* Cambridge: The M.I.T. Press.

-- 1995. “The Philosophical Significance of Church’s Slingshot.” *Mind*, New Series, Vol. 104, No. 416 (Oct., 1995), 761-825.
--2001. *Facing the Facts.* Oxford: Clarendon Press.

Olson, Ken. 1987. *An Essay on Facts.* Palo Alto: CSLI Publications.

Putnam, Hilary. 1962. "The analytic and the synthetic" In Herbert Feigl and Grover Maxwell, eds., *Minnesota Studies in the Philosophy of Science*, III, Minneapolis: University of Minnesota Press.

Quine, Willard van Orman. 1990/1951. “Two Dogmas of Empiricism.” In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.
--1953. “Reference and Modality.” In *From a Logical Point of View.* Cambridge: Harvard University Press.
--1960. *Word and Object.* Cambridge: The M.I.T. Press.
-- 1966a. “Three Grades of Modal Involvement.” In *The Ways of Paradox.* Cambridge: Harvard University Press.
--1966b. “Vagaries of Definition”. In *The Ways of Paradox.* Cambridge: Harvard University Press.

Russell, Bertrand. 1990/1905. “On Denoting”. In A.P. Martinich (ed) *The Philosophy of Language* Oxford: Oxford University Press.
-- 1957/1910. “Knowledge by Acquaintance and Knowledge by Description". In *Mysticism and Logic.* Garden City, New Jersay: Doubleday.
-- 1919. *Introduction to Mathematical Philosophy.* London: George Allen Unwin.
--1921. *The Analysis of Mind.* London: George Allen Unwin.

Stalnaker, Robert. 1991. “Propositions”. In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.

Strawson, Peter. 1990/1950. ‘On Referring’ in A.P. Martinich (ed) *The Philosophy of Language.* Oxford: Oxford University Press.

Tughendat. Ernst. 1970. “The Meaning of ‘Bedeutung’ in Frege”. In *Analysis*., xxx (1970), pp. 177-189.

Weisler, Stephen. 1991. “An Overview of Montague Semantics.” *Meaning and Truth.* 1991. Jay L. Garfield and Murray Kitely (eds.) St. Paul, Minnesota: Paragon House

Whitehead, Alfred North and Russell, Bertrand. (1997/1912). *Principia Mathematica (to *56).* Cambridge: Cambridge University Press.

Sentential Reference the Foundations of Semantics

Introduction

Frege held that sentences refer to truth-values. His argument for this is known as the Slingshot. The purpose of this work is to respond to Frege's bold view.

In outline, my response is this. First of all, we must distinguish sentence-types from sentence-tokens. The position that "sentences" refer is ambiguous between the position that sentence-types refer, on the one hand, and the position that sentence-tokens refer, on the other.

Uncontroversially, sentence-types don't refer to anything. The *type* "I am tired" is not associated with any one proposition or, therefore, any one truth-value.

Sentence-*tokens* do not refer to truth-values. The relationship between a token of "I am tired" and the truth-value *true* is not even remotely like the relationship between a token of "Plato" and Plato.

The Slingshot involves two major blunders. It involves an equivocation on the notion of "logical equivalence". More importantly, it involves an erroneous conception of reference.

A correct conception of reference leads to an astonishing result. Where expression-*tokens* are concerned, there is no difference between reference and semantic contribution – no difference between *meaning and reference.* Expressions that *seem* not to refer to anything (e.g. tokens of "and", "not" and morphemes like case endings) are plausibly seen as referring to higher-order functions.

One important consequence of this – by no means the only one – is that *sentence-tokens refer to their meanings.* In part, though not entirely, the meaning of a sentence-token is a proposition. So part of what a sentence-token refers to is the propositions it bears. But the meaning of a sentence-token is not *just* the proposition it encodes. After all, "Plato snores" and

"does Plato snore?" encode the very same proposition, but palpably have different meanings. After we take into account semantic differences like the one just mentioned, it turns out that the meaning of a sentence-token is a function involving, but not identical with, a proposition. So it turns out that sentence-tokens refer to such functions.

It is the purpose of this work to show that the claims just made, though controversial, have deep roots in some orthodoxies of contemporary semantics. Indeed, if we are to avoid adopting a picture like the one just described, we must jettison some deeply entrenched and extremely plausible, if not truistic, semantic views.

Chapter 1 Do Sentences Refer to Truth-Values?

Frege famously held that sentences refer to *truth-values.* Frege's position is by no means universally accepted. Some hold that sentences do not refer at all, while others, e.g. Barwise and Perry, hold that sentences refer not to truth-values but to propositions. I hope to make clear that this entire debate needs to re-examined.

Types versus Tokens

Uncontroversially, sentence-*types* don't refer to anything. The question is: What, if anything, do sentence-tokens refer to?

Let us start with some foundational points. The sentence-*type* "I am tired" has a meaning, of course, but its meaning is not a proposition. Its meaning is, very roughly, a function by which to derive propositions from the contexts of utterance. The meaning of such a *token*, then, is a proposition.[237] If Bob

says “I am tired” at 3:30 p.m., his utterance is true just in case Bob is tired at 3:30 p.m. If Mary says “I am tired” at 4:40 a.m., her utterance is true just in case Mary is tired at 4:40 a.m. The proposition affirmed by Bob has different truth-conditions, and is therefore distinct, from that affirmed by Mary. So while the *type* “I am tired” is not itself true or false, and thus doesn’t encode a proposition, nonetheless it *assigns* propositions to its tokens on the basis of the circumstances under which those tokens are produced.

Indexicals are a feature of nearly every sentence in a natural language. But what about those sentence-types that lack an indexical-component? Perhaps, it might be suggested, such sentence-types have propositions, not functions, for their meanings. There are, however, two problems with this proposal.

First, as noted, indexical sentences are the rule, not the exception. Note that all, or almost all, sentences of natural language contain tense-markers and thus indexicals.[238] Sentences containing indexicals have a “two-dimensional” structure. There is what indexical sentence-*types* mean (functions that assign propositions to sentence-tokens). And there is what indexical sentence-tokens mean (propositions). So, uncontroversially, *some* sentences have a two-dimensional structure.

Second, even if, in natural language, there *were* non-indexical sentence-types, it would be very strange to suppose that *some* sentence-types had propositions for their meanings, while others had functions from contexts of utterance to propositions for their meanings. Such a supposition would introduce a massive cleavage into a class that, in all likelihood, is unified. In fact, arguably, it would amount to a redefinition of “sentence”. There would be “sentences$_1$” – expressions that have functions from contexts of utterance to propositions for their meanings. This would comprise all, or nearly all, of the natural-language expressions we refer to as “sentences”. But there would

also be "sentences$_2$": these would be sentence-types that had propositions for their meanings. Sentences$_1$ and sentences$_2$ are obviously very different sorts of things, and it is hard to believe that the class of "sentences" is so heterogeneous – that "I am tired" is in a different category from "two plus two equals four".

Of course, in some cases, the tense-marking is obviously irrelevant. If I say "2 plus 2 equals 4", the tense-marker isn't doing any work. But that doesn't mean that *semantically* it is not an indexical. We know, on independent grounds, that considerations of time-indexing are irrelevant in this kind of case. So we ignore the tense-marker – we ignore the indexical.

We must distinguish matters of *semantics* from matters of epistemology. It is very hard to believe that *semantically* the tense-marker in "John is in the kitchen" is different from that in "two is prime". We can adapt an argument of Kripke's (1977) to substantiate this. That argument, adapted, is: If we *were* dealing with a case of semantic ambiguity, we'd expect there to be languages that had *different* words or morphemes: one for the time-innocuous tense-marker in "2 plus 2 equals 4", and another for the non-innocuous one in "John is in the kitchen". I doubt there are such languages. So we are probably not dealing with semantic ambiguity.

The natural solution is to take sentence-types as uniformly having functions for their meanings: functions from contexts of utterance to propositions.[239] Although in some cases it might *seem* as though the sentence-type *per se* has a proposition for its meaning, that is, I believe, an illusion. No one would maintain that the sentence-*type* "I am tired" refers to truth-values or propositions. That type is not associated with any one proposition or, therefore, with any one truth-value. So it doesn't refer to truth-values or to propositions. By way of conclusion, given that all natural language sentences are in the same class as "I am tired", it is clear that

sentence-types don't refer to anything. The *type* "I am tired" doesn't refer to anything. Its semantic content is a function (from contexts of utterance to propositions). This point generalizes to all, or most, other sentence-types. Granting this, the next question is: what might *tokens* of sentences refer to?

The debate regarding sentence-reference must be reconstructed within this framework. Once such a two-dimensionalism about sentences is accepted, analogical reasoning leads us say that sentence *tokens* refer to propositions. As we noted, the *type* "you" doesn't refer to anything. Its semantic content is a function (from contexts of utterance to individuals). However, the semantic content of a *token* of "you" is some individual. The semantic content of such a token coincides with its referent.

The sentential analogue of this point is as follows. The type "I am tired" doesn't refer to anything. Its semantic content is a function from contexts to propositions. The semantic content a token of "I am tired" is the proposition assigned to it by that function. The semantic content of a *token* of "I am tired" coincides with its referent. Once we acknowledge the two-dimensional nature of sentences, analogical reasoning suggests that sentence-tokens refer to their *meanings* – to the things they are assigned by the corresponding types. Those things are not truth-values.

Of course, analogies are not always probative. Fortunately, we can do better than an analogy.

Some Platitudes about Reference

When we examine the concept of reference, the idea that sentences (or, more precisely, sentence-tokens) refer to truth-values is put under severe strain.

Suppose you want to make a statement about Plato. What you'll do is

use some expression that *refers* to him, e.g. “Plato” or perhaps “the author of the Republic”. Suppose you want to ascribe wisdom to Plato. You will say “Plato was wise” or perhaps “the author of the Republic was wise”. There is obviously an essential connection between the fact that “Plato” refers to Plato and the fact that, if you want make a statement about Plato, you can use the word “Plato” (in a purely literal manner[240]) to do so.

Suppose that “Plato was wise” meant *Charlemagne was wise.* Suppose that, in general, sentences of the form “…Plato…” meant:…*Charlemagne*… In that case, it would be very hard to maintain that “Plato” referred to Plato; and it would, to say the least, be natural to suppose that “Plato” referred to Charlemagne.

Suppose you want to make a statement about the truth-value *true.* What you’ll do is use some expression like: “the truth-value *true*” or “the property of being true”. Perhaps you’ll use some neologism like “**T**”. But what you *won’t* do is use “snow is white” or “grass is green”. Those expressions would be absolutely useless; you couldn’t possibly use them to talk about truth-values, except (at most) in some circuitous sense.

Finally, consider the sentence:

(*) “the last thing Plato said had the truth-value *true*.”

If, as Frege and Church maintain, “snow is white” really refers to the truth-value true, then we should be able to use that sentence-token to refer to that truth-value. We should thus be able to rephrase (*), using “snow is white” instead of the “truth-value true”. But this doesn’t seem remotely possible. If you want to produce a sentence that means: *everything Bob says has the truth-value true*, you’ll say “everything Bob says has the truth-value *true*” or

"everything Bob says is true". You *won't* say "everything Bob says has snow is white".

Let us consolidate these points. Sentence-tokens of the form "…snow is white…" don't mean:…*the truth-value* true… "Snow is white and grass is green" doesn't mean: *the truth-value* true *and grass is green.* "Possibly, snow is white" doesn't mean: *possibly, the truth-value* true. By contrast, "Plato is wise" means *Plato is wise.* In general, sentence-tokens of the form "… Plato…" mean:…*Plato.*…As we've seen, this is indissolubly connected with the fact that "Plato" refers to Plato. So the relationship between a token of "snow is white" and the truth-value *true* isn't remotely like the relationship between "Plato" and Plato.

Clearly the idea that "snow is white" refers to the truth-value *true* is under serious strain. If "snow is white' really refers to the truth-value *true*, why is it so hard to make it act as though it did? Why is it so hard to make it behave like expressions that we *know*, for certain, to refer to that truth-value?

Of course, one can say:

Semantically 'snow is white' is synonymous with 'the truth-value *true*'. The differences between them that you've just described are pragmatic.

But this obviously over-extends the concept of pragmatics. Pragmatics cannot be used to exonerate *every* semantic theory. Suppose I say that "grass is green" means: *penguins don't fly.* My theory will conflict with what "grass is green" is usually taken to mean, and that conflict will obviously have a semantic basis.

Suppose I dug in my heels and said:

Semantically my theory is right. Admittedly, the facts appear to be

discrepant with my theory. But that has to do with pragmatics, not semantics.

Clearly I would be over-using the concept of pragmatics. The problem is that "penguins don't fly" and "snow is white" are *too* remote in meaning. "Snow is white" *seems* no less remote in meaning from "the truth-value *true*". Arguably, they are *more* remote in meaning than the first pair: the first pair consists of two indicative sentences; the second consists of a sentence and sub-sentential expression. So it is hard to see how pragmatics could be used to exonerate the view that those two expressions co-refer.

Here we would do well to consider a point made by Max Black. Replacing a referring term with a co-referring term will not turn a meaningful utterance into a meaningless one. "The author of the Republic was wise" is meaningful. "The author of the Parmenides" co-refers with "the author of the Republic." Therefore "The author of the Parmenides was wise" is meaningful. Given this, let us suppose, with Frege, that "snow is white" co-refers with "the True". Obviously "if three has no factors, then three is prime" is meaningful. If Frege is right, then replacing the antecedent with "the True" ought to be meaningful. But it is not: "if the True, then three is prime" is nonsense.

Black's point works in the other direction. Consider a token of: "everything Plato said has the property of truth". If you replace "the property of truth" with "snow is white" what results is nonsense: "everything Plato said has the property of snow is white".

Alonzo Church rejects Black's point:

A more serious error, in the reviewer's opinion, is the fallacious attempt to refute Frege's view, that sentences are designations of truth-values, by reference to the grammar of English language. It is pointed out that if the sentence "Three is a prime" is a designation of the True (Black uses a capital

letter in translating Frege's *das Wahre*), then the expressions "Three is a prime" and "the True" ought to be interchangeable in a non-oblique contexts – as indeed Frege himself maintained in analogous cases. Thus from "If three is prime then three has no factors" we get "If the True then three has no factors." To the latter expression Black objects calling it nonsense, and by saying that it "has no more use than" the expression "if seven then three has no factors." In the absence of supporting reasons for such objections, it must be supposed that Black is rejecting the expression in question on the ground that it violates the rules of English grammar. But surely the right question to ask here is not what, by existing custom, are the rules of English, but rather what it is desirable to take as the rules of a formalized language. In a suitable formalized language the analogue of "if the True then three has no factors" does have a use, namely as a designation of the False. And indeed it is not unusual, in formulations of propositional calculus, and of formalized languages containing propositional calculus, to introduce primitive constants denoting one or both of the two truth-values, and to allow substitution of such constants (as well as of longer sentences) for the propositional variables. [241]

Church seems to be saying (among other things) that the meaninglessness of "if the True, then three is prime" is a mere fact about English grammar. But that is not so. The meaninglessness of that expression reflects deep logical facts. Those facts are, in their turn, reflected in facts about grammaticality. But we are not dealing with *mere* ungrammaticality.

Consider a case of a sentence-token that can plausibly be regarded as *merely* ungrammatical:

(1) "John no like play soccer".

(2) is ill-formed. But it is clear enough what it means or, at any rate, what it is supposed to mean: *John doesn't like to play soccer*.

But

(2) "if the True, then three is prime"

is in a very different category. The connective "if…then…" expresses a relation between truth-evaluable entities (sentence-tokens or propositions); it expresses the relation of consequence or entailment. The True is not itself true or false; neither is "the True"; therefore those things entail nothing and nothing entails them. Only propositions (or, if you are a nominalist, sentences-tokens) stand in entailment relations. The True is not such a thing. The meaninglessness of (2) obviously reflects that fact. That is not a peculiarity of grammar. It is a deep fact about the concept of entailment.

Church talks about the "analogue" of (2) in formal languages. If by "analogue" Church means "translation", then Church's point is spurious. Translations preserve meaning – or the lack thereof. Any correct translation of (2) will be as meaningless as (2). So Black's point will apply to any translation no less than to (2).

Of course, we could use (2) to mean:

(3) *given any true proposition P, P entails that three has no factors*.

Perhaps when Church talks about "analogues" of (2), he is referring to sentences that mean (3). But when (2) expresses a proposition like (3), "the True" doesn't function as an expression that refers to the property of truth. It functions rather as a quantifier. (Thus, depending on one's views on

quantifiers, it is functioning either as something that doesn't refer at all or as something that refers to a second-level function.)

The difference between "snow is white" and "the True" isn't remotely like the difference between "I" and "me". In the one case, we are dealing with something that is plausibly seen as a mere grammatical difference. In other case, we are dealing with a much deeper difference: a difference that is reflected in, but not confined to, a grammatical difference.

There is a subtler point to make here. Arguably, grammatical differences *are* semantic differences. There are grammatical differences between "red" and "John". You can say "John smokes", but not "red smokes". ("The property of redness smokes" is false, perhaps even absurd, but perfectly grammatical. But "red smokes" is simply ungrammatical.) But these grammatical differences presumably reflect semantic differences. It could even be maintained (in fact, I will maintain it later) that these semantic differences reduce to differences in what the expressions in question *pick out.* The idea would be that the one picks an individual, while the other picks out a function.[242]

To take another example: There are profound grammatical differences between "or" and "Smith". But those differences obviously reflect semantic differences. And a case *can* be made (I will try to make it) that those semantic differences reduce to differences in what those expressions pick out. ("Or" picks out a higher-order function, "Smith" picks out an individual.)

It is by no means clear that grammatical facts can ever be dismissed as *merely* grammatical – as semantically innocent. On the face of it, they seem to embody semantic facts; and grammatical categories seem to be *semantic* categories. When you say that "John" and "some man" are "noun phrases", you are presumably making a statement about the semantic roles of such expressions. By this reasoning, when you say that "I" and "me" are

grammatically different, you are saying something about their semantic roles.

The temptation is to say that “I” and “me”, and “eager” and “eagerly”, differ *merely* in ways that are grammatical. But that is, I believe, quite false. I would suggest that there are subtle semantic differences between (say) “I” and “me”. The function of a token of “I” is not *merely* to contribute an individual. That is merely a good first approximation. It would be very strange if it turned out that deeply entrenched grammatical facts – like the difference between “I” and “me” – were semantically innocent.

Obviously when semanticists produce semantic models, they have to abstract from subtleties – just as physicists must initially idealize away from wind-resistance and the like. But we must not confuse the model with the thing modeled. In some of our models of English, there is no place for a difference between “but” and “and”. But obviously there are differences between those expressions. In some of our models of English, there is no difference between “I” and “me”. But that doesn’t mean that there *is* no semantic difference. On the face of it, there almost certainly is. In any case, what I am saying is that we cannot assume that grammatical facts are of no semantic consequence. It is quite possible that grammatical differences *do* reflect semantic differences. In that case, Black’s point prevails *even if* the difference between “snow is white” and “the True” is “merely” grammatical.

Other Reasons to Question Frege’s View

The semantic content of a token of “you” is the individual assigned to that token by the function encoded in the corresponding type. The semantic content of a token of “I am tired” is the proposition assigned to that token by the function encoded in the corresponding type. In general, the semantic content of an expression-token is the thing assigned to that token by the

corresponding type.

The semantic content of a token of “I am tired” is some proposition[243]; it is not the truth-value *true.* The function encoded in the type “I am tired” assigns truth-evaluable entities to the relevant tokens; it assigns things that are true or false. Propositions are true or false. The truth-value *true* is not itself true or false. The semantic content of a (non-abortive) token of “I am tired” or “John is in Cleveland” is something that is true or false. The truth-value *true* is not itself true or false; so it is not the semantic content of any such token.

There are additional reasons to doubt that sentence-tokens refer to truth-values. Consider the sentence-token “George W. Bush smoked a cigarette on April 2, 2005”. You and I both understand that expression as well as it can be understood. We completely and thoroughly grasp its semantics. But we don’t have the slightest idea whether it is true or not. Except perhaps in very special cases, knowledge of the semantics of a sentence-token absolutely does not require knowledge of its truth-value.[244]

If a term refers, its referent is an integral part of its semantic content, and you cannot possibly know its semantics unless you know what it refers to. There are, of course, *apparent* counter-examples to this: you can know the meaning of “the number n such that n=1 if Goldbach’s conjecture is true and n=0 otherwise. But, as we’ll see soon enough, this view involves a confusion of type-semantics with token-semantics.

To illustrate my general view, let me start with a tendentious example. Suppose that “Blark” is somebody’s name, but you have no idea whose name it is. Because you don’t know who “Blark” refers to, you really don’t know its semantics. You don’t quite know what is meant by sentence-tokens like “Blark is a nice person”; you don’t know what proposition they encode. Of

course, you know that such sentence-tokens will be true just in case somebody named “Blark” is a nice person. But the proposition encoded in “Blark is a nice person” is not: *somebody named ‘Blark’ is a nice person.* So if a term refers, one cannot know its semantics without knowing its referent.

Let us apply this point to sentence-tokens. I know the *meaning* – the semantic content -- of a token of “George Bush is currently napping”. But I don’t know its truth-value. In general, one can know exactly what a sentence-token means – one can know exactly what its semantic content is -- without knowing its truth-value. As we just saw, if a term refers, one cannot possibly have a complete knowledge of its semantic content unless one knows to what it refers. One *can*, however, have a complete knowledge of the semantics of a token of “George Bush is currently napping”. If such a token *did* refer to a truth-value one would have to know that truth-value to have a complete grasp of its semantics. It follows that such a token does *not* refer to a truth-value.

To understand a sentence token, one has to know what *would* make it true, not *whether* it is true. Indeed, language would have little point if one knew the truth-value of every sentence-token one understood. So the truth-value of a sentence-token is not part of its semantic content. It follows that *sentence-tokens don’t refer to truth-values.*

An obvious response this line of reasoning takes this form:

It is obviously very reasonable to suppose that definite descriptions are singular terms – that “the inventor of bifocals” refers to Benjamin Franklin and “the even prime” refers to the number two. Given this, consider the definite description “the richest man in Holland”. You can understand it perfectly without having the foggiest idea to whom it refers. Consider the expressions: “the chemical-compound that cures AIDS”, “the sequence of sentences that proves the most important mathematical result of the 23rd

century", "the exact number of people in Nepal". You have no idea what these expressions refer to. But you understand them perfectly. So one *can* grasp the semantics of an expression without knowing its referent. So given only that one can grasp the semantics of a sentence without knowing its truth-value, it does not follow that a sentence's referent is anything other than its truth-value.

This is an important objection. But it is based on a failure (once again) to distinguish type-semantics from token-semantics, and is completely fallacious.

For the sake of argument, let us suppose, with the objector, that definite descriptions are in fact singular terms (as opposed to Russellian quantifiers). Let w be our world and let w* be some other world where English is spoken. So "the inventor of bifocals" has the same *semantics* in both w and w*. Further, suppose that, in w*, Thomas Jefferson (as opposed to Benjamin Franklin) invented bifocals.

We have a problem. By supposition, "the inventor of bifocals" has the same *semantics* in both w and w*. Presumably, the referent of a referring term is at least part of its semantic content. ("Plato" refers to Plato, and that fact is essential to the semantics of "Plato".) "The inventor of bifocals" refers to one thing in w and to a different thing in w*. Thus, it would appear, "the inventor of bifocals" does *not* have quite the same semantic content in w* that it has in w. So "the inventor of bifocals" both does, and does not, have the same semantics in w and w*.[245] A paradox.

There is an easy way to make sense of all this. Strawson[246] showed us how. Let us suppose, with the objector, that definite descriptions are singular terms. We begin by distinguishing the expression-*type* "the inventor

of bifocals" from *tokens* of that expression. The expression-*type* doesn't refer to anything. But the expression-type does have meaning – it does have semantic content. The meaning of that expression-type is a rule that assigns referents to *tokens* of it, depending on the context of utterance. (In this case, the world of utterance is at least part of the context of utterance.)

So the meaning of the expression-*type* "the inventor of bifocals" is some rule like the following: Given any token of "the inventor of bifocals", if that token occurs in a context where there is exactly one unique bifocal inventor x, then that token refers to x. [247]

Thus, in w, *tokens* of "the inventor of bifocals" refer to one thing (Benjamin Franklin), while in w*, such tokens refer to a different thing (Thomas Jefferson). So the semantic content of a *token* of "the inventor of bifocals" in w is different from the semantic content of a *token* of that expression in w*. Accordingly if, as the objector assumes, definite descriptions are singular terms, there is no one entity that has the same semantics in different worlds while having different referents. The meaning of the expression-*type* "the inventor of bifocals" is the same in both w and w*. But the expression *type* doesn't refer to anything. So, trivially, it doesn't refer to different things in different worlds. *Tokens* of that type are what refer. In w, such tokens refer to Benjamin Franklin. In w*, such tokens refer to Thomas Jefferson.

The principle I am defending is this: If an expression refers to something, it is not possible to have complete knowledge of its semantics without knowing its referent. Definite descriptions are not counter-examples to this principle. Definite description-*types* don't refer to anything. So, vacuously, they are not counter-examples to the main thesis. Tokens of such types do indeed refer. But you cannot possibly know the semantic content of such a token if you don't know what it refers to. The semantic content of a

token of “the inventor of bifocals” in w is Franklin. If you are in w and you don’t know what a given token of that expression refers to, then you simply don’t know its semantic content. So, once again, the principle is upheld.

Here is an analogy – I think it is more than an analogy. “You” obviously refers to different people on different occasions. But it doesn’t undergo a change in *meaning* every time it is used to refer to a new person. It is not as though “you” is ambiguous between Fred and Charlie and Mary and Joanne. “You” is not like “bank”.

Here we have two options. One is to say that “you” has a variable reference, but a constant meaning: its referent changes, but not its semantic content. But we’ve already seen why that would be a tortured and arbitrary position.

This brings us to the other option. The *type* “you” always has the same semantic content. That content is a function from contexts of utterance to individuals. A given *token* of you has such an individual for its content. The type “you” doesn’t refer to anything. So that type is vacuously consistent with our principle that, *if* an expression refers, then one must know to what it refers in order to have a complete knowledge of its semantics. Tokens of the type “you” do indeed refer. The semantic content of such a token coincides with its referent. So there is plainly no way to know the semantic content of such a token unless one knows to what that token refers. Thus the principle is again upheld.

To sum up, whenever we have something that refers, it turns out that one *does* have to know its referent to know its semantics. Apparent examples to the contrary are based on a failure to distinguish type-semantics from token-semantics. One can completely understand a token of “George Bush is napping” without knowing its truth-value. Therefore that token doesn’t refer to a truth-value.[248]

Another Objection

Here is an objection to our view on sentence-reference:

Some connectives, e.g. “and”, are truth-functional. “snow is white and grass is green” is true iff both “snow is white” and “grass is green” are true. The truth of the conjunction is a function of the truth-values of its components. So surely the truth-values of “snow is white” and “grass is green” are at least *among* the things which they semantically contribute to “snow is white and grass is green”. In general, it seems that, in some contexts, what a sentence semantically contributes is a truth-value. So the thesis that sentences don’t semantically contribute their truth-values is wrong.

This reasoning is quite spurious. Consider the sentence-token:

(kg) “Kermit is green”.

Whether (kg) is true is a function of whether Kermit is green. But the semantic contribution of “Kermit” is not the property of being green: “Kermit” semantically contributes Kermit, not the property of being green. What contributes the property of being green is the expression “green” (or “is green”). So even though the truth of (kg) *depends* on whether Kermit is green, “Kermit” obviously doesn’t contribute that property to the meaning of (kg).

Now consider the sentence-token:

(js) “John believes that snow is white”.

Whether (js) is true is a function of whether a certain proposition has a certain property: it is a function of whether the proposition *that snow is white* has the property of being believed by John. Plainly “that snow is white” doesn’t semantically contribute the property of being believed by John. What contributes that property is the phrase “is believed by John”.

Now consider the sentence-token:

(sg) “snow is white and grass is green”.

(sg) comprises tokens of “snow is white” and “grass is green”. Whether (sg) is true *depends* on whether the propositions encoded in those sentence-tokens have a certain property (the property of being true), just as the truth of “Kermit is green” depends on whether Kermit has a certain property. But it doesn’t follow that those constituent sentence-tokens *contribute* that property. Demonstrably, they do not. The sentences themselves no more contribute the property of truth than the occurrence of “Kermit” in (kg) contributes the property of being green and “that snow is white” in (js) contributes the property of being believed by John. In (sg), what contributes the property of truth is the “and”, which can be thought of as a second-level predicate – as denoting a property of propositions or sequences of propositions.

(sg) is equivalent to

(sg*) and*<snow is white, grass is green>,

where “and*” is a second-level predicate: one that ascribes truth to the

propositions encoded in the sentence-tokens falling in its scope.

So the “and” in (sg) can be thought of as ascribing truth to both members of that sequence.[249] Thus, the reason that “snow is white” occurs truth-functionally in (sg) is *not* that it contributes the property of truth; it is that the property of truth is *ascribed* to the proposition it contributes.

These points are worth looking at from a slightly different angle. The truth of (kg) is a function of whether the thing meant by “Kermit” has the property of being green. This is equivalent to saying: (kg) *attributes* the property of being green to the thing meant by “Kermit”. So the semantic role of “Kermit” is *not* to contribute that property; it is to contribute the thing to which that property is being attributed.

The truth of (sg) is a function of whether the things meant by “snow is white” and “grass is green” are true. This is equivalent to saying: (sg) *attributes* truth to those things. So the semantic role of “snow is white” and “grass is green” is *not* to contribute that property. It is to contribute the things to which that property is being ascribed.

(sg) is truth-functional because a certain property is being *ascribed* to the propositions encoded in the constituent sentence-tokens, *not* because those sentences themselves contribute that property. A truth-functional occurrence of a sentence-token is one in which truth is *attributed* to the corresponding proposition, not one in which that sentence itself contributes the property of truth.[250]

Chapter 2 The Slingshot

So far we’ve seen some reasons to hold that neither sentence-types nor

sentence-tokens refer to truth-values. But there is Frege's famous argument to the contrary, dubbed the Slingshot by Barwise and Perry (1975). The term is meant to convey how several modest and seemingly truistic assumptions sustain a quite formidable semantic thesis. I wish to show that the Slingshot is not to be trusted, for it proves things that nobody would accept as true, no matter what their semantic views are.

First of all, what is the Slingshot? As Frege presented this thesis, it asserts that, if a sentence refers at all, its referent is its corresponding truth-value. To illustrate: Let S be any sentence-token, and let P be any proposition. Starting with S, we can produce a sentence-token that has P for its meaning. All we have to do is replace referring terms with co-referring terms. Even if we are limited to that one operation, we can still reach our destination.

Consider the sentence-token "Socrates was wise." Suppose we want to produce a sentence-token that has for at least part of its meaning the proposition: *Mozart wrote fewer fugues than Bach.* And suppose we only allow ourselves to replace referring terms with co-referring terms.

We start with:

(i) "Socrates was wise."

We replace "Socrates" with a co-referring term:

(ii) "The unique thing x such that [Mozart wrote fewer fugues than Bach and x is identical with Socrates] was wise."

(ii) appears to have, for at least part of its meaning, the proposition: *Mozart wrote fewer fugues than Mozart.*

Replacing a referring term with a *co*-referring term surely doesn't result in

a change of *reference*. Let us refer to this principle as (CR): co-referring terms preserve referentiality.

Given (CR), (i) and (ii) must co-refer. At the same time, (i) and (ii) have different propositions for their meanings. (ii) is about Mozart (among other things), while (i) is not about Mozart. Whatever it is that a sentence-token refers to, it must be invariant with respect to replacements of referring terms, in that sentence-token, with co-referring terms. Propositions are *not* thus invariant. So sentence-tokens don't refer to propositions.

What is thus invariant is a sentence-token's *truth*-value. Replacing a referring term with a co-referring term never turns truth into a falsity or falsity into truth. So sentence-tokens, if they refer to anything, refer to truth-values.

Following Church (1943[251]), we can make this argument tighter.

Church 1943

Let S and S* be any two non-analytic sentence-tokens that have the same truth-value.

(1) S
(2) The class of all things x such that (x=x and S) is identical with the class of all things x such that (x=x).
(3) The class of all things x such that (x=x and S*) is identical with the class of all things x such that (x=x).
(4) S*.

Let us assume that:

(LL) logically equivalent expressions co-refer.

(1) and (2) are logically equivalent. So (1) and (2) co-refer. (3) is what results when a referring term in (2) is replaced with a co-referring term. So (2) and (3) co-refer. (3) and (4) are logically equivalent. So, by (LL), (3) and (4) co-refer. Thus, (1) and (4) co-refer. What we've just said about (1) and (4) can be said (mutatis mutandis) about any two sentences that are like in truth-value. So any two sentence-tokens, alike in truth-value, co-refer.

If we accept only these assumptions, we can prove some very perplexing things. For example, Davidson[252] "proves" that there is only one fact!

Pick any two true sentence-tokens, say: "Grass is green" and "Whales are mammals". Davidson's argument shows that those two sentences are made true by the very same fact:

(i) The fact that: grass is green.
(ii) The fact that: the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x).
(iii) The fact that: (x=x and whales are mammals) is identical with the class of all things x such that (x=x).
(iv) the fact that: whales are mammals.

(i) and (ii) must co-refer. For given (LL), "grass is green" must refer to the same thing as "the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x)". So, given (CR), (i) and (ii) corefer.

(ii) and (iii) co-refer for the same reason, and (iii) and (iv) also co-refer for

that very reason. So “the fact that grass is green” refers to the same thing as “the fact that whales are mammals”. Accepting this conclusion, Davidson reasons that there is only one fact. He used this fact (pun intended) as a way of justifying some far-reaching epistemological views.[253]

Here is an analogue of Davidson’s argument.

(I) The proposition: that grass is green.
(II) The proposition that: the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x).
(III) The class of all things x such that (x=x and whales are mammals) is identical with the class of all things x such that (x=x).
(IV) The proposition that: whales are mammals.

(I) and (II) must co-refer. For given (LL), “grass is green” must refer to the same thing as “the class of all things x such that (x=x and grass is green) is identical with the class of all things x such that (x=x)”. So, given (CR), (I) and (II) corefer. (II) and (III), and also (III) and (IV), co-refer for the same reason. So “the proposition that grass is green” refers to the same thing as “the proposition that whales are mammals”. Therefore the proposition that whales are mammals is identical with the proposition that grass is green. Any two true propositions are identical. Thus, there is only one true proposition. Obviously there are many true propositions. So the Slingshot argument must rest on a fallacy.

There is an even more counterintuitive application of the Slingshot argument. Some philosophers distinguish between constitution and identity. The statue is not *identical* with the clay because you can destroy the statue

without destroying the clay. But the clay obviously *constitutes* the statue.

Everyone agrees that every physical entity is either identical with, *or* constituted by, displacements of mass-energy. Your car is *constituted* by a long, complicated series of mass-energy displacements. Let S be a sentence that describes exactly that series. Your favorite shirt is constituted by some entirely distinct series of mass-energy displacements. Let S* be a sentence that describes exactly that series.

For obvious reasons, the series of mass-energy displacements that makes S true is identical with the series that makes S, conjoined with "triangles have three sides", be true.

(A) The set comprising all and only the mass-energy displacements that make true the sentence: S.

(B) The set comprising all and only the mass-energy displacements that make true the sentence: The class of all objects x such that (x is identical with x and S) is identical with the class of all objects x such that (x is identical with x)

(C) The set comprising all and only the mass-energy displacements that make true the sentence: The class of all objects x such that (x is identical with x and S*) is identical with the class of all objects x such that (x is identical with x)

(D) The set comprising all and only the mass-energy displacements that make true the sentence: S*.

The set comprising all and only the mass-energy displacements that make up your favorite shirt has exactly the same members as the set of such displacements that make up your car. Any two objects are constituted by exactly the same mass-energy displacements. Given any two physical objects,

they are either identical, or one of them constitutes the other. This is what the Slingshot proves.

Here is a similar argument. Obviously different physical objects are involved in different events. One set of objects was involved in Kennedy's assassination (a certain bullet, a certain car, and so on). Another was involved in Mozart's composition of the Jupiter Symphony (a certain quill pen, sheets of paper, and so on).

When I say "involved", I mean directly involved; I am not talking about remote causal involvement. So the bullet that killed Kennedy was directly involved in that man's assassination. But the chopping down of the tree that yielded the wood composing Oswald's rifle was not directly involved.

Of course, analytic sentences do not describe events at all. So "Kennedy was assassinated" and "Kennedy was assassinated and 1+1=2" describe the same event. The second encodes information not encoded in the first. But that information has nothing to do with *events*; it has to do with timeless relations among numbers.

The core of the argument is given by the following sequence of expressions:

(a) The set of physical objects involved in the event described by the following sentence: Kennedy was assassinated.

(b) The set of physical objects involved in the event described by the following sentence: The class of all objects x such that (x is identical with x and Kennedy was assassinated) is identical with the class of all objects x such that (x is identical with x)

(c) The set of physical objects involved in the event described by the following sentence: The class of all objects x such that (x is identical with x and Mozart composed the Jupiter symphony) is identical with the class of all

objects x such that (x is identical with x)

(d) The set of physical objects involved in the event described by the following sentence: Mozart composed the Jupiter symphony.

(a) and (b) co-refer by LL and CR. (b) and (c) co-refer by CR and LL. (c) and (d) co-refer by LL and CR

So if we accept the assumptions that drive the Slingshot, we must accept that (a) and (d) co-refer. The set of physical objects directly involved in Kennedy's assassination is identical with the set involved in Mozart's composing of the Jupiter symphony. The set of physical objects directly involved in *any* event is identical with the set of physical objects directly involved in any other.

This is obviously false. Also, it seems to imply that there is only one physical object. Presumably x and y are *different* physical objects iff they are involved in different events. So there is only one physical object. This is what the Slingshot proves.

Chapter 3 The Problems with the Slingshot

The Slingshot proves things that aren't true. So we should look for problems with it. If I am not mistaken, there are at least two. The first is identified by Barwise and Perry (1983).

Consider the following two sentence-tokens:

(i) Socrates was wise.

(ii) The unique thing x such that Mozart wrote fewer fugues than Bach and x is Socrates, was wise.

There are two ways to read the definite description in (ii). It can be read as a term that refers to individuals (as a "singular term"), or it can be read as a quantifier.

Referring terms semantically contribute their referents, and nothing else. Given this, suppose we read the definite description in (ii) as a term that refers to an individual. In that case, (ii) has exactly the following proposition for its meaning:

(ii_R) Socrates was wise.

If we treat definite descriptions as expressions whose tokens refer to individuals, then (ii) and (i) have *precisely* the same meaning. Given that a referring term semantically contributes its referent and nothing else, it follows that co-referring terms contribute the exact same thing. Thus, replacing referring terms with co-referring terms cannot change meaning.

Replacing referring terms with referring terms not only preserves truth-value: it preserves *meaning.* Since CR preserves truth-value *and* meaning, we have no grounds for saying that a sentence-token refers to anything *other* than its meaning.

Suppose we treat definite descriptions as quantifiers. Does that make things better for Frege's argument? It does not.

If definite descriptions are quantifiers, then (ii) means:

(ii_{RUS}) For some x, x was uniquely such that [Mozart wrote fewer fugues than Bach and x was Socrates] and x was wise.

Russell thought that quantifiers, and thus definite descriptions, don't refer to anything. So if Russell is right, then (ii) is not what results when a referring term in (i) is replaced with a co-referring term. If definite descriptions don't refer, then Frege's argument crashes: CR doesn't guarantee the needed steps.

To be thorough, we must consider a nuance. A case can be made that Russell's analysis of definite descriptions actually *is* consistent with the idea that they are referring terms. If he is right, they are, at least arguably, terms that refer to second-level functions.[254] Let C be the class of wise objects. (ii_{RUS}) is true exactly if C contains exactly one object x such that (x is identical with Socrates and Mozart wrote fewer fugues than Bach), and no such object falls outside of C. So if Russell's analysis is correct, the definite description in (ii) can be seen as a function that assigns truth to a class exactly if that class contains a unique x such that (x is identical with Socrates and Mozart wrote fewer fugues than Bach), and no such object falls outside of that class.

If definite descriptions do refer to functions, then (ii) is indeed what results when a referring term in (i) is replaced with another referring term. *But the two terms do not co-refer.* For "Socrates" refers to an individual, and therefore doesn't co-refer with anything that refers to a function.

Let us sum up. If definite descriptions refer to individuals, then (i) and (ii) mean exactly the same thing. If definite descriptions are quantifiers, then (ii) is not what results when a referring term in (i) is replaced with a co-referring term. In either case, Frege's argument is destroyed.

These considerations break Church 1943. (Let Omni be the class of all self-identical objects.) If we read the definite descriptions as expressions that refer to individuals, we have:

(1_R) S

(2_R) Omni=Omni.

(3_R) Omni=Omni.

(4_R) S*.

There is no logical equivalence between the first two steps, or the last two. The argument fails.

So if Church 1943 is to work, the definite descriptions must be read as quantifiers. But then there is no logical equivalence between any two of the entries. Let S be the sentence-token "Mozart wrote fewer fugues than Bach" and let S* be the sentence-token "whales are mammals".

In that case, (2_R) and (3_R) become existence claims that are not logically equivalent. For even though Omni is identical with Omni in all worlds, there are logically possible worlds where whales are mammals and where Mozart wrote more fugues than Bach. (2_R) is contingent on the truth of *Mozart wrote fewer fugues than Bach.* (3_R) is not contingent on that truth. So (2_R) and (3_R) are not logically equivalent.

Let us sum up. There are two ways to read the descriptions in (1)-(4): as singular terms or as quantifiers. If they are read as singular terms, then there is no logical equivalence between the first two entries or between the last two: and the argument fails. If the descriptions are read as quantifiers, then (3) is not what results when a referring term in (2) is replaced with a *co*-referring term: and the argument fails. Basically, as Barwise and Perry observed, once we are clear on how we read definite descriptions, the Slingshot crumbles.[255]

Frege's way out

Frege and Church wouldn't necessarily be moved by the argument just given; for they wouldn't agree with the conception of reference that underlies it. In this section, I would like to show why their disagreement is innocuous, at least as far as the cogency of our argument is concerned.

Here is the controversial assumption that I made. *If* (e.g.) "Plato" and "the unique x such that 1+1=2 and x=Plato" co-refer, then they both contribute exactly the same thing, namely Plato. Consequently, replacing the one with the other will *not* change sentence-meaning; and this, as we saw, breaks the Slingshot. (If replacing the one with the other *does* change sentence-meaning, that is because they are not *co*-referring terms: one of them is a singular term that refers to a philosopher while the other is a quantifier.)

Let us generalize this. I assumed that if E is an expression that refers to some object O, then when E occurs in a sentence-token, its semantic contribution to the sense of that token is O: nothing more, nothing less. Of course, *if* this is right, *and* E and E* both refer to O, then they both contribute exactly the same thing, namely O.

But Frege (and Church) wouldn't agree with the conception of reference that underlies our attack on his argument. He says that referring terms have both "sense" and reference.[256] A referring term semantically contributes not (only) its referent but its *sense*. Co-referring terms can have different senses. So it *is* possible to change sentence-meaning by replacing referring terms with co-referring terms.

Let us make this more concrete. In Frege's view, "the teacher of Aristotle" and "the author of the *Republic*" are both referring terms that refers to Plato. So they co-refer. But according to Frege, they make different semantic contributions. When "the teacher of Aristotle" occurs in a sentence,

it contributes some concept to the proposition encoded in that sentence. When “the author of the *Republic*” occurs in a sentence, it contributes a different concept. That is why

(*) “the teacher of Aristotle was wise”

encodes a different proposition from:

(**) “the author of the *Republic* was wise”.

Let us evaluate this view. Consider the expression “the teacher of Aristotle”. What is the concept contributed by that expression? Here a slight detour may be useful.

Consider the sentences (or sentence-tokens):

(i) “A teacher of Aristotle was wise.”
(ii) “Three teachers of Aristotle were wise.”
(iii) “No teachers of Aristotle were wise.”

Each makes a statement about the concept: *teacher of Aristotle.* (i) says that something instantiating that concept was wise. (ii) says that three things instantiating that concept were wise. (iii) says that nothing instantiating that concept was wise.

Clearly, all of “a teacher of Aristotle”, “no teachers of Aristotle”, and “three teachers of Aristotle” semantically contribute the concept: *teacher of Aristotle.* Perhaps they make other contributions. But that concept is their *common* contribution.

So *if* "the teacher of Aristotle" contributes a concept, it contributes (inter alia) that concept (*teacher of Aristotle*). Thus, if Frege is right, (*) encodes a proposition that has that concept as one of its constituents; it thus makes a statement about that very concept. In this respect, it is just like (i)-(iii).

Supposing this is right, what is that statement? What does (*) say about the concept: *teacher of Aristotle*? If the concept *teacher of Aristotle* is uninstantiated, then (*) will be false. If there are *multiple* (equally contextually salient) instances of that concept, (*) will be false. So that concept must be uniquely instantiated. Further, that instance must be wise.

So *if* (*) encodes a proposition concerning the concept *teacher of Aristotle*, then (*) is true exactly if that concept is uniquely instantiated and any instance of it is wise. [257]

Even referentialists (those who think that definite descriptions are singular terms) agree that (*) is true exactly under that circumstance.[258] Strawson agrees with Russell that "the king of France is bald" is true just in case there is at least one king of France, at most one of king of France, and any king of France is bald. Russell says that if there is no such object, then the statement is false; Strawson says that if there is no such object, then the statement is abortive. They both agree that it is true exactly if those three conditions are met.

So everybody – referentialists and descriptivists alike -- agrees that *if* "the greatest teacher of Aristotle" contributes a concept or sense, then (*) means:

($*_R$) Exactly one thing was a greatest teacher of Aristotle and any such thing was wise.

But if ($*_R$) gives the meaning of (*), then "the teacher of Aristotle"

doesn’t refer to Plato at all. It is a quantifier. So either it doesn’t refer to anything or it refers to a function. If we say that “the teacher of Aristotle” contributes a concept, then we must embrace Russell’s theory. The essence of that theory is: definite descriptions do not refer to individuals; they are quantifiers.

By exactly similar reasoning, if we say “the author of the *Republic*” contributes a sense, we are saying that it is a quantifier. In that case, it either refers to a function or it doesn’t refer to anything.

If “the author of the *Republic*” and “the teacher of Aristotle” contribute different senses, it is because they *don’t* co-refer. If, as Russell thought, they don’t refer to anything, then they don’t *co*-refer. If they refer to functions, they refer to different functions, as we saw earlier. If those descriptions are quantifiers, they are different quantifiers, and thus don’t co-refer.

A corollary is that replacing a definite description with a co-referring definite description cannot result in a change in meaning. Replacing “the author of the *Republic*” with “the teacher of Aristotle” results in a change of meaning only if those expressions are entirely different quantifiers, and thus either denote *different* entities (different functions) or denote nothing at all. So there is no way to preserve meaning by replacing a definite description with a co-referring definite description.

Admittedly, our argument is not yet airtight. For, conceivably, one could say that “the teacher of Aristotle” semantically contributes *both* Plato *and* the concept *teacher of Aristotle*. But this position is a non-starter. If it is right, then “the teacher of Aristotle is wise” encodes a proposition that concerns that concept *and* that concerns Plato. If “the teacher of Aristotle is wise” encodes a proposition about the concept *teacher of Aristotle*, it means: *exactly one thing was a teacher of Aristotle and any such thing was wise.* If “the teacher of Aristotle is wise” encodes a proposition about Plato, it means:

Plato was wise. If it does *both*, then it must presumably mean: *Plato was wise and exactly one thing was a teacher of Aristotle and any such thing was wise*. But that is obviously not what is meant by "the teacher of Aristotle is wise". The very idea that it has that meaning is too tortured to consider.

There is more specific point to make. If definite descriptions were *both* quantifiers *and* terms that refer to individuals, then "the teacher of Plato" and "the author of the *Republic*" would not, strictly speaking, be co-referring terms. Up to a point, they would co-refer. But past that point, they would either not be referring terms, and thus not *co*-referring terms, or they would refer to different things. In either case, it would be false to say that, on the whole, those descriptions were co-referring terms. Replacing "the teacher of Plato" with "the author of the *Republic*" would indeed result in a change of meaning, but only because those two expressions would *not* be co-referring terms. Replacing "the inventor of bifocals" with "the first post-master general", or "Plato" with "the unique thing x such that arithmetic is incomplete and x=Plato", would result in a change of meaning, but only because the pairs in questions didn't really co-refer. CR would not permit the substitution of, say, "Plato" with "the unique thing x such that arithmetic is incomplete and x=Plato". So CR would no longer permit exactly the substitutions that the Slingshot depends on.

Nobody denies that "the inventor of bifocals" is semantically associated with the concept *inventor of bifocals*, whereas "the first post-master general" is *not* semantically associated with that concept. But semantic association is not always semantic contribution. There are two ways that the concept *inventor of bifocals* can figure in the semantics of "the inventor of bifocals". On the one hand, that concept can *itself* be (part of) what is contributed by "the inventor of bifocals" to sentence-tokens in which that expression occurs. One the other hand, the semantic role of that concept can be to *pick out* the

thing which that expression semantically contributes to such tokens. Strawson took the second view, Russell took the first.

Once again, it might help to think about demonstratives. Obviously "you" is semantically associated with a concept (something like *addressee of these words*). The same is true of every indexical. "Yesterday" is semantically associated with *day before this day.* "That man" is associated with *contextually salient man.* And so on.

But when you say "you are tired" to Fred, your proposition is true exactly if *Fred* is tired. You are not *saying* anything about the concept *addressee of these words.* You are making a statement about a person, not a concept. You are not saying that something x is an addressee of these words and that any such addressee is tired. You are saying that *Fred* is tired. This doesn't mean that the concept *addressee of these words* is semantically irrelevant: it is highly relevant. But its relevance is confined to *picking out* the thing to which the property of being tired is ascribed. That concept does not *itself* figure in the proposition affirmed: it serves merely to *identify* or *fix* the identity of that proposition (or, more precisely, of some constituent of it).[259]

Let us map this point onto definite description. *If* definite descriptions are singular terms, then a token of "the inventor of bifocals snored" attributes the property of being a snorer to *Franklin*. Obviously the concept *inventor of bifocals* is relevant. But its relevance is *not* that it is itself a constituent of the proposition affirmed; its relevance is that it serves to *pick out* the thing to which the property of being a snorer is ascribed.

The temptation is to say that *inventor of bifocals* is a constituent of the proposition expressed. Very well. That is a perfectly plausible position. But if we take that line, then "the inventor of bifocals" becomes a quantifier. Why? If "the inventor of bifocals" contributes the concept *inventor of bifocals* to sentence-tokens in which it occurs, then a token of "the inventor of bifocals

snored" encodes something of the form:…*inventor of bifocals*…The question is: what, in that case, does "the inventor of bifocals snored" *say* about that concept? Everybody – friends and foes of Russells' theory – agrees that *if* "the inventor of bifocals snores" is to be true, it is necessary and sufficient that there be at least one (contextually salient) bifocal inventor, at most one such inventor, and that any such inventor be a snorer. So *if* "the inventor of bifocals snored" encodes a proposition about the concept *inventor of bifocals*, then it says that exactly one thing instantiates that concept and that any such thing snores. In that case, "the inventor of bifocals" turns into a Russellian quantifier.

So *if* "the inventor of bifocals" contributes a concept, then it is a quantifier. If it is a referring term, then it does not contribute that concept: that concept, in that case, serves the function of *picking out* the thing which that expression *does* contribute to sentence-tokens in which it occurs. If a concept is contributed, the description is a quantifier: and the Slingshot fails, as we saw. If the concept is *not* contributed, then the Slingshot fails, as we saw.

A Salmon-Kripke reply

I must consider a response to my argument (due to Nathan Salmon):

Here is a response to your argument. Never mind English definite descriptions. Instead we introduce an artificial device "ze" that works according to the following rule: *ze phi*[260] refers to the unique individual that satisfies (has) phi, if there is such an individual, and fails to refer otherwise. In that case "Scott" and "ze author of Waverly" are co-referential. Now replace "the" through Church's (or Frege's) argument with "ze". The resulting argument is immune to what you've said.

Let us see how well this proposal fairs.

(i) Sir Walter Scott is Sir Walter Scott.
(ii) Sir Walter Scott is ze author of Waverly.
(iii) Sir Walter Scott is ze man who wrote twenty-nine Waverly novels altogether.
(iv) Ze number n, such that Sir Walter Scott wrote n Waverly novels, is twenty-nine.
(v) Ze number of counties in Utah is twenty-nine.

Of course, “ze author of Waverly” contributes the very same thing as “Scott”. They both contribute Scott. So the proposition encoded in (ii) is identical with the proposition encoded in (i).

“Ze man who wrote twenty-nine Waverly novels altogether” contributes Scott. So (iii) encodes the same proposition (ii) and as (i).

“Ze number n, such that Sir Walter Scott wrote n Waverly novels” refers to/contributes the number twenty-nine. So (iv) encodes the proposition: twenty-nine is twenty-nine.

“Ze number of counties in Utah” refers to/contributes the number twenty-nine. So (v) encodes the proposition: twenty-nine is twenty-nine.

So when we rewrite Church’s argument, replacing “the” with “ze”, and consider the propositions encoded in the resulting sentences, the result is this:

(i) Scott is Scott.

(ii) Scott is Scott.
(iii) Scott is Scott.
(iv) Twenty-nine is twenty-nine.
(v) Twenty-nine is twenty-nine.

(iii) and (iv) are not synonymous, or "very nearly so". So the argument fails.

Let us refer to expressions of the form "ze phi" as "zefinite descriptions". By our stipulation, "ze unique x such that 1+1=2 and x=Plato" semantically contributes Plato and Plato alone. So "Plato snores" is synonymous with "ze unique x such that (1+1=2 and x=Plato) snores" and with "ze unique x such that (arithmetic is incomplete and x=Plato) smokes." This blocks Frege's argument.

In general, replacing a referring term with a co-referring zefinite-description preserves meaning. Replacing definite descriptions with zefinite descriptions will not help any slingshot; it will actually ensure that meaning *is* preserved and will prevent the divergence of meaning from truth-value that slingshots depend on.

So our argument against the Slingshot has nothing to do with the *actual* semantics of definite descriptions. Replace definite descriptions with zefinite descriptions. The result will be the same – the Slingshot will still fail.

Chapter 4 The Second Problem with the Slingshot

Some Slingshot-style arguments involve the assumption that logically equivalent *sentences* (or, more exactly, sentence-tokens) co-refer. Church

1943 is an example of such an argument. So are Gödel 1944, Quine 1953, and Davidson 1967. Unless it is assumed that logically equivalent sentence-tokens co-refer, these arguments immediately break down. For those arguments explicitly rely on that assumption to generate the needed steps. [261]

The principle that logically equivalent sentences (sentence-tokens) co-refer is thought to be an innocuous extension of the principle that logically equivalent *sub-sentential* expressions (e.g. “Plato” and “the unique thing x such that 1+1=2 and x-=Plato”) co-refer.

But it is not an innocuous extension of that principle. It is an entirely different principle, and it is almost certainly false.

Let us start with some foundational points. Earlier we argued that sentence-*tokens* are what encode propositions. (Sentence-types encode functions that assign propositions to their tokens.) If this is right, then we shouldn’t say that “sentences” are logically equivalent; we must talk about sentence-*tokens* instead.[262]

Given this, what is it for two sentence-tokens to be logically equivalent? Let T and N be sentence-tokens (not types):

(T) “Triangles have three sides.”

(N) “1+1=2”.

T and N are “logically equivalent”. Each of those tokens semantically contributes a proposition. T and N are logically equivalent in the sense that the one proposition entails the other. For two sentence-tokens to be logically equivalent is for the proposition semantically contributed by the one to entail the proposition semantically contributed by the other.

Now consider the expressions:

(P) “Plato”

and

(P*) “the unique x such that 1+1`=2 and x=Plato”.

P and P* are sub-sentential expressions. They are “logically equivalent”. But what does this mean? It does *not* mean that the things they semantically contribute are logically equivalent. The thing they semantically contribute is Plato. Plato is not a proposition. So he is not logically equivalent with anything, not even himself.

When we say that P and P* are “logically equivalent”, what we mean is: for all values of x, the condition, i.e. the propositional function that a thing must satisfy to be referred to by P entails and is entailed by the condition a thing must satisfy to be picked out by P*. Permit me to elucidate this.

Given the semantic rules of English, the condition that a thing must satisfy to be referred to by P is:

(*) x is identical with Plato.

The corresponding condition for P* is:

(**) x is such that 1+1=2 and x=Plato.

For all values of x, (*) entails and is entailed by (**).

Let us hearken back to T and N. When we say that T and N are “logically

equivalent" we are *not* saying that, for all values of x, the propositional function that a thing must satisfy to be referred to by T entails and is entailed by the condition a thing must satisfy to be picked out by N.

When we say that T and N are "logically equivalent", we *are* saying that the propositions they semantically contribute are logically equivalent. We are talking about a relationship that holds among the things semantically contributed by those expressions.

But when we say that "Plato" and "the unique x such that 1+1=2 and x=Plato" are logically equivalent, we are *not* making a statement about the things contributed by those expressions. Those expressions contribute an individual; and an individual isn't logically equivalent with anything.

So the sense in which T and N are "logically equivalent" is entirely different from the sense in which noun-phrases, like P and P*, are logically equivalent.

Thus, given only that "logically equivalent" *noun* phrases – e.g. "Plato" and "the unique thing x such that 1+1=2 and x is identical with Plato" -- must co-refer, it does not follow that "logically equivalent" sentence-tokens must co-refer. The term "logically equivalent" denotes completely different relations in the two cases.

In fact, if two expressions are logically equivalent in the one sense, that practically guarantees that they *won't* be logically equivalent in the other sense. Logically equivalent sentence-tokens typically do *not* semantically contribute the same thing. A token of "1+1=2" contributes one proposition; a token of "triangles have three sides" contributes a very different proposition. By contrast, logically equivalent noun-phrases necessarily *do* contribute the very same thing. "Plato" and "the unique x such that 1+1=2 and x=Plato" necessarily *do* contribute the very same thing, namely Plato. Logically equivalent noun-phrases *do* necessarily contribute the same thing. Logically

equivalent sentence-tokens typically do *not* do so. So if two expressions are "logically equivalent" in the one sense, that practically guarantees that they *won't* be "logically equivalent" in the other. "1+1=2' and "triangles have three sides" are *not* logically equivalent in the sense in which "Plato" and "the unique x such that x=Plato" are logically equivalent.

The only sentence-tokens that *are* logically equivalent in the latter sense are those that encode the exact same proposition, i.e. are those that are synonymous, like "snow is white" and "schnee ist weiss". The kind of "logical equivalence" that guarantees co-reference is the kind that "logically equivalent" sentences *don't* typically have.

Slingshot advocates assume that the kind of "logical equivalence" that obtains between P and P* is identical with the kind that obtains between "1+1=2" and "triangles have three sides". That assumption is wrong. Without it, many Slingshots fail.[263]

Consider Davidson's argument:

Let S and S* be any two sentences that have the same truth-value.

(1) S
(2) The class of all things x such that (x=x and S) is identical with the class of all things x such that (x=x).
(3) The class of all things x such that (x=x and S*) is identical with the class of all things x such that (x=x).
(4) S*.

According to Davidson, this sequence of sentence (sentence-tokens) shows that S and S* must co-refer. To generate this result, Davidson needs that (1) and (2) corefer, and that (3) and (4) co-refer. He says that they do co-refer;

and his grounds are that “logically equivalent” sentences co-refer. But this principle is spurious, as we saw.

There is more to say on this matter. Right now, I can only make some anticipatory remarks.

A case can be made that reference is indistinguishable from semantic contribution. Let me clarify my meaning. “Plato” refers to Plato exactly if sentences of the form “…Plato…” bear propositions of the form:…*Plato*.… This suggests that for “Plato” to refer to Plato just *is* for Plato to be the thing semantically contributed by tokens of “Plato” to the meanings of sentence-tokens in which that expression occurs. This, in turn, suggests that semantic contribution and reference are the same thing.

Sentence-tokens semantically contribute propositions. Tokens of “snow is white” semantically contribute *snow is white*. The same is true *mutatis mutandis* of any other sentence-token. This suggests that sentence-tokens *refer* to the propositions that are their meanings. If this is right, then logically equivalent sentences *always* refer to different things, except in the special case where those sentence-tokens are *synonymous*. (This means that, *within a single language*, it will be very hard to find *different* sentence-tokens that co-refer. Perhaps “Smith is an enemy” and “Smith is a foe” co-refer. Even that is dubious.) If this is right, then “logically equivalent” sentences that encode different propositions unfailing *don’t* co-refer.

Obviously these are not innocuous ideas. But they are defensible: and we will defend them. One has to bite some very hard bullets to escape them – one must say that “Plato” can refer to Plato without semantically contributing Plato. Many Slingshots involve the idea that logically equivalent sentences that encode *different* propositions refer to the same thing. That idea, we will find, is dead wrong. Such pairs necessarily *don’t* co-refer, at least if some exceedingly plausible principles about reference are granted.

Chapter 5 Some Famous Slingshots

In this chapter, I will discuss why, in light of the points just made, some famous applications of the Slingshot cannot be regarded as cogent.

Church's Slingshot

In his *Introduction to semantics*, Carnap argued that sentences refer to propositions. Alonzo Church supposedly showed that, given principles that Carnap accepts, sentences must refer to truth-values. One of those principles is: logically equivalent sentences co-refer.

Church's argument is this. Let S and S* be any true, non-analytic sentences.

(i) S
(ii) The class of all things x such that (x=x and not S) is identical with the null class.
(iii) The class of all things x such that (x=x and not S*) is identical with the null class.
(iv) S*.

(i) and (ii) are "logically equivalent"; so, by (LL), they must co-refer. (iii) is what results when a referring term in (ii) is replaced with a co-referring term; so by (CR), (ii) and (iii) must co-refer. (iii) and (iv) are "logically equivalent", and must therefore co-refer.

Let us examine this argument. The definite descriptions in (ii) and (iii) are either singular terms or they aren't. Let us consider each case. First, suppose they are not singular terms. That means they are Russellian quantifiers. In that case, (ii) and (iii) become existence claims that are *not* logically equivalent. Let S be the sentence "Mozart wrote fewer fugues than Bach" and let S* be the sentence "Bill Clinton was president for two terms". There are logically possible worlds where Bill Clinton was not President for two terms and where Mozart wrote fewer fugues than Bach. So, on that interpretation, (ii) and (iii) are not logically equivalent.

Let us now suppose that definite descriptions are singular terms. In that case, they both pick out some entity. That entity is a certain class. Let NULL be that class. Remember that, if E refers to O, then O is E's sole semantic contribution to the meanings of sentence-tokens in which E occurs. So, if the definite descriptions in (ii) and (iii) are singular terms, then (ii) is synonymous with:

(ii$_S$) NULL is identical with NULL.

For the same reasons, (iii) also becomes synonymous with (ii$_S$). So if the definite descriptions in Church's argument are treated as singular terms, then that argument amounts to this:

(i) S
(ii$_S$) NULL is identical with NULL.
(iii$_S$) NULL is identical with NULL.
(iv) S*.

(i) and (ii_S) are not logically equivalent; neither are (iii_S) and (iv). So the argument fails.

Regardless of whether we read the description in Church's argument as singular terms or as quantifiers, the needed pairs of sentences are not logically equivalent, and the argument therefore fails.

But even if we *grant* that the needed pairs are logically equivalent, the argument still fails. "Plato" and the "the unique x such that 1+1=2 and x is Plato" are "logically equivalent" in the sense that, given the semantic rules of English, they necessarily contribute the same thing. The thing they contribute isn't logically equivalent with anything. "1+1=2" and "triangles have three sides are "logically equivalent" in a completely different sense. The thing *contributed* by the one entails the thing contributed by the other. Given the semantic rules of English, the thing contributed by the one is necessarily *not* identical with the thing contributed by the other. The sense in which those two sentences are logically equivalent has nothing to do with co-reference. So logically equivalent sentences *don't* typically co-refer, contrary to what Church's (1943) argument needs.

Quine's Slingshot

Let us now discuss Quine's astonishing application of the Slingshot to the nature of modality.[264] Some sentence-connectives are "extensional". The truth of "snow is white and grass is green" is wholly a function of the truth-values of the sentences falling within the scope of the "and". So "and" is extensional. Some sentences very much appear *not* to be extensional. The truth of "Khrushchev resigned after Kennedy died" is *not* a function of the truth-values of the constituent sentences ("Kennedy died", "Khrushchev resigned"). Both those sentences are true; but that is compatible with

"Khrushchev resigned after Kennedy died" being either true or false.

Orthography doesn't always make it clear what the *scope* of a sentence connective is. In the sentence "Grass is green and snow is white", the *scope* of the "and" is really the (ordered) pair <grass is green, snow is white>.[265] English-notation doesn't make this fact clear. (So-called "Polish" logical notation does make it clear.) The "and" can be thought of as a function that assigns truth to a sequence of expressions exactly if both of them express true propositions. But what is important is that *both* sentences fall within the scope of the "and".

The same thing is true of "after". In the sentence "Khrushchev resigned after Kennedy died", *both* constituent sentences fall within the scope of the "after", even though only one of those sentences is to the right of it.

Two more pieces of terminology must be defined. Let us say that a connective is PSLE[266] iff there is no change in truth value as a result of replacing a sentence falling in its scope with a logically equivalent sentence. A clear example of such a connective is "necessarily".

Finally, let us say that a sentence-connective "induces opacity" iff sentences falling with its scope do not have their ordinary meaning. "And" does not induce opacity; for "snow is white" has the same meaning in "snow is white and grass is green" that it has on its own. I cannot give an uncontroversial example of a connective that induces opacity. I myself don't think that *any* sentence-connectives induce opacity. That is actually one of the things I will try to show. But Frege thought that epistemic connectives – e.g. "John believes that" -- induced opacity.

A Slingshot-style argument can be constructed that shows that at least *one* of the following two propositions is true:

(TF) All PSLE sentence-connectives are truth-functional.
(OP) All PSLE sentence-connectives induce opacity.

Let “Phi” be an arbitrary PSLE connective. Let S and S* be any two sentences that have the same truth-value.

TF

(1) Phi (S)”.
(2) Phi (the class of all things x such that (x is identical with x and S) is identical with the class of all things x such that (x is identical with x)).
(3) Phi the class of all things x such that (x is identical with x and S*) is identical with the class of all things x such that (x is identical with x)).
(4) Phi (S*).

Given that “Phi” is PSLE, we clearly have safe passage from (i)-(iv). It seems to follow that *any* PSLE connective is truth-functional.

But there are two reasons why Quine’s argument fails. Trivially, either definite descriptions are singular terms or they aren’t. If they aren’t, they’re Russellian quantifiers. Let’s consider each case.

If they are Russellian quantifiers, then TF fails miserably. The transition from (2) to (3) no longer goes through. Let S be the sentence “Bach wrote fugues”. Let S* be the sentence “Stalin was evil”. In that case, (2) and (3) become:

(2_R) There is some class C such that, for any x, x falls into C iff (x=x and

Bach wrote fugues), and there is some class C* such that, for any y, y falls into C* iff (y=y), and C=C*.

(3_R) There is some class C such that, for any x, x falls into C iff (x=x and Stalin was evil), and there is some class C* such that, for any y, y falls into C* iff (y=y), and C=C*.

(2_R) and (3_R) are definitely *not* logically equivalent. So the fact that "Phi" is PSLE does *not* warrant the transition from (2) to (3), at least not if definite descriptions are read as Russellian quantifiers.

One might say:

Obviously (2_R) and (3_R) are not logically equivalent. But this doesn't mean that the transition from (2) to (3) fails. It means that it fails *only* if "Phi" is given wide-scope, i.e. only if (2_R) and (3_R) in their entirety fall within the scope of "Phi". But if "Phi" is given narrow-scope, then the transition from (2_R) to (3_R) works. There are different narrow-scope readings. Perhaps TF goes through on one of them.

Pick any narrow-scope reading you wish. The result will *in*validate one of the steps in TF. Here is the most obvious narrow scope reading of (2) and (3).

(2_R) There is some class C such that, for any x, x falls into C iff (x=x and Bach wrote fugues), and there is some class C* such that, for any y, y falls into C* iff (y=y), and Phi (C=C*).

(3_R) There is some class C such that, for any x, x falls into C iff (x=x and

Stalin was evil), and there is some class C* such that, for any y, y falls into C* iff (y=y), and Phi (C=C*).

Let us put this reading into the context of the argument as a whole:

(1) Phi (Bach wrote fugues)
(2_R) There is some class C such that, for any x, x falls into C iff (x=x and Bach wrote fugues), and there is some class C* such that, for any y, y falls into C* iff (y=y), and Phi (C=C*).
(3_R) There is some class C such that, for any x, x falls into C iff (x=x and Stalin was evil), and there is some class C* such that, for any y, y falls into C* iff (y=y), and Phi (C=C*).
(4) Phi (Stalin was evil).

In that case, the steps from (1) to (2), and from (3) to (4), become completely invalid. For any objects C and C*, the proposition *Bach wrote fugues* is not equivalent to the proposition *C=C**. The same is true of *Stalin was evil.* So if we give "Phi" that particular narrow scope reading, then two much needed transitions simply fail; they are no longer guaranteed by the fact that "Phi" is PSLE.

Granted, there are other narrow-scope readings. But it is easy enough to verify that, on any of them, one of the steps in question becomes invalid. It would be a tedious and unnecessary exercise to go through each one of them.

It is clear what we must say. We must treat the definite descriptions involved as expressions that refer to individuals. If we *don't* do this, then TF becomes a tissue of non-sequiturs.

But TF becomes a tissue of non-sequiturs if we *do* treat definite descriptions in that way. For the sake of argument, let the definite

descriptions in TF be referring expressions, Let Omni be the class of all things x such that x=x. In that case, all of the definite descriptions refer to Omni.

Consequently, TF becomes:

(1) Phi (Bach wrote fugues).
(2) Phi (Omni=Omni).
(3) Phi (Omni=Omni).
(4) Phi (Stalin was evil).

In that case, TF is patently bankrupt. For no two of the following sentences are logically equivalent: “Bach wrote fugues”, “Omni=Omni”, “Stalin was evil”. Consequently, the fact that “Phi” is PSLE doesn’t guarantee *any* of the steps in TF.

Quine’s Slingshot fails for another reason. It uses the principle that logically equivalent *sentences* co-refer. As we’ve seen, the kind of logical equivalence that holds between sentences has nothing to do with co-reference. So the needed principle is false.

Gödel’s Slingshot

According to Gödel[267], the Slingshot shows *either* that there is one fact *or* that Russell’s theory of descriptions is correct. Given this ugly choice, Gödel rightly infers that Russell’s theory is correct.

But there is no need to make that choice. For Gödel’s Slingshot involves the muddle we’ve been discussing.

Let “Fa” and “Gb” be any two true, non-analytic sentences. Assume that a is not identical with b. Finally, assume that the meaning of a referring term is

the thing it refers to.

1. Fa. *premise*
2. a is the unique thing x such that [Fx and x=a]. *corefers with 1 by LL*
3. a is the unique thing x such that [a is not b and x=a]. *corefers with 1 by LL and CR*
4. a is not identical with b. *corefers with 3 by LL*
5. b is the unique thing y such that [y is not identical with a and y=b]. *corefers with 4 by LL*
6. b s the unique thing y such that [Gy and y=b]. *corefers with 5 by CR*
7. Gb *corefers with 6 by LL.*

If cogent, Gödel's argument shows that *either* definite descriptions must parse out, in the way described by Russell's theory, *or* that any two true sentences have the same meaning. If Russell's theory is right, then CR cannot be used to generate the needed transitions. Obviously not all sentences have the same meaning. So Russell's theory must be right.

There are two major problems with Gödel's argument. First, that argument relies heavily on the idea that logically equivalent sentences co-refer. We've already seen why this assumption is unwarranted and probably false.

Let us discuss the second problem with Gödel's theory. This one is identified by Barwise and Perry (1983). Tokens of definite descriptions are either quantifiers or singular terms. Let us consider each case.

First let us suppose that tokens of definite descriptions refer to individuals. Remember what we established earlier: if two terms co-refer, they make exactly the same semantic contribution. The following two expressions co-refer:

(i) “a”
(i) “the unique thing x such that [Fx and x=a]”
(iii) “the unique thing x such that [a is not b and x=a].”

So they both contribute a and a alone.
The following three expressions all co-refer.

(iv) “b”
(v) “the unique thing y such that [y is not identical with a and y=b]”.
(vi) “the unique thing y such that [Gy and y=b].”

So they all contribute b and b alone.
Thus Gödel’s argument becomes:

1. Fa. *premise*
2. a is a. *corefers with 1 by LL*
3. a is a. *corefers with 1 by LL and CR*
4. a is not identical with b. *corefers with 3 by LL*
5. b is b. *corefers with 4 by LL*
6. b is b. *corefers with 5 by CR*
7. Gb *corefers with 6 by LL.*

But this is an obvious failure. “Fa” is not logically equivalent with “a is a”. (Remember that “Fa” is non-analytic.) So even if logically equivalent sentences *did* co-refer, (1) and (2) would not co-refer. For the same reason, neither would (6) and (7). So the argument is a failure if definite descriptions refer to individuals.

It is a failure if they don't refer to individuals. If they don't refer to individuals, they are quantifiers. If they are quantifiers, then either they don't refer to anything or they refer to functions. If they don't refer to anything, then (CR) no longer permits the steps from (2) to (3) or from (5) to (6).

If definite descriptions refer to functions, then CR *still* cannot be used to make those steps. If they refer to functions, then "the unique thing x such that [Fx and x=a]" refers to a very different function from "the unique thing x such that [a is not b and x=a]." There are possible worlds (this is not one of them) where English is spoken where those functions assign different truth-values to the same entity.

On any reading of definite descriptions, Gödel's argument fails.

There is yet another reason why Gödel's argument fails. Replace each occurrence of "the" in that argument with "ze", as defined earlier. Let Gödel$_Z$ be the resulting argument. Gödel$_Z$ is an exact analogue of Gödel's original argument. Nothing can *prove* that "ze phi" is not a referring term; for, by our stipulation, it is just that. But *if* Gödel$_Z$ is cogent, then it "proves" that "ze phi" doesn't refer. If Gödel$_Z$ is cogent, it isn't cogent; therefore it isn't cogent. Therefore, the same is true of Gödel's original argument.[268]

Chapter 6 Extensions and Intensions

Carnap (1947, 1958) said that that the intension of a "sentence" (he didn't distinguish tokens from types) is a proposition, and that the extension of a sentence is a truth-value.

This view is generally seen as either true or, if not quite true, then very reasonable: a reasonable extension of the term "extension".

Carnap's view, I am afraid, doesn't deserve quite such a warm response.

Let us start with a more straightforward case of an expression that is alleged to have both extension and intension.

The word “whale”, we are told, has both intension and extension. The intension is a property or function (from objects to the truth-values). The extension is the class of whales (Shamu, Willy, Moby Dick…)

But there is no one entity that has both intension and extension. Tokens of “whale” have extension; the expression-type “whale” has intension. Nothing has both.

We don’t want to say that the expression-type “whale” denotes *different* classes in different worlds, while having the same semantics. That is perilously close to absurdity. The *extension* of a term is the set of objects that it refers to. So unless we divorce reference from semantic content, it is self-contradictory to say that “whale” has the same semantics in worlds where it has different extensions.

What we must say is: *tokens* of “whales” denote different classes in different worlds; the *type* has an invariant meaning: it pairs off a token of “whale” in a world W with a class C iff C is the class containing all and only whales in W.

The type “whale” has an intension. That intension of “whale” is a function from worlds to classes: a function that assigns truth to a class C in a world w exactly if C comprises all the whales in w and nothing that is not a whale.

Tokens of “whale” have *extensions.* In this world (at this time), it has as its extension the class containing all and only: *Shamu*, *Willy*…

By parity of reasoning, sentence-types have *intension*; sentence-tokens have extension.

The intension of a sentence is some kind of function. But it cannot be a function from worlds to truth-values, for the reasons we’ve seen. The type “I am tired’ does not bear any proposition; so it cannot have a proposition for its

intension. The same must be true of the type “Socrates was wise”. If we assign a different kind of intension to the type “Socrates was wise”, we end up committed to the highly arbitrary view that sentences can have entirely different sorts of things for the intensions: the intension of “Socrates was wise” is propositions (functions from worlds to truth-values). The intension of “I am tired” isn’t a proposition at all – it is a function from contexts of utterance to propositions.

In so far as the type “I am tired” has a meaning, it is a function from contexts of utterance to propositions. So if anything is its intension, it is that function. If anything is the intension of “Socrates was wise”, it is a function from contexts (worlds and possibly times) of utterance to propositions.

The *extension* of a token of “I am tired” must be the thing that the *intension* of the corresponding type assigns to that token – just as the *extension* of a token of “whale” is the thing that the intension of the corresponding type assigns to that token. So the extension of a token of “I am tired” is the *meaning* of that token. Thus, the extension of a token of “Socrates was wise” is the thing it means (roughly, the proposition *that Socrates was wise*); it is not, *pace* Carnap, a truth-value.

A *proposition* can be seen as a function from worlds to truth-values. So, at least arguably, the extension of *Socrates is wise* is the truth-value *true* in worlds where he is wise, and the truth-value *false* in worlds where he is not.

But the extension of a token of “I am tired” is not a truth-value. Its extension is the thing which the function associated with the corresponding type assigns to that token. That thing is, very roughly, a proposition. That proposition *in its turn* has a truth-value. But the truth-value is not the thing assigned to that token by the corresponding type. So Carnap’s view is simply wrong.

Also, I don’t think it is quite correct to say that *propositions* have

extensions. Intensions and extensions are things that expressions have. Propositions aren't expressions. Strictly speaking, it is no more accurate to say that *Socrates is wise* has the "extension" *true* for this world as an argument than it is to say that "F(x)=2x" has the "extension" four for two as an argument.

Before he did his work on indexicals, Kaplan[269] once wrote:

[T]he question of the truth value of which a given proposition actually is a concept and the individual of which a given individual concept actually is a concept are…empirical. That is, although a given proposition may actually be a concept of Truth, there are usually other possible states with respect to which it is a concept of Falsehood.

A proposition is indeed a concept. But it is a concept of a world, not of a truth-value. A concept of Socrates is some concept C such that Socrates falls under C. *x is a greatest teacher of Plato* is a concept of Socrates because Socrates falls under that concept. Given that Socrates falls under that concept, we can say that it assigns *truth* to Socrates. A concept is a function from objects to truth-values. *x is a greatest teacher of Plato* is not a concept of the truth-value that it assigns to Socrates; it is a concept of Socrates.

If a proposition is a concept, it is a function that assigns truth-values to objects. If a proposition is such a function, the objects in question are worlds. Indeed, propositions are often analyzed as functions that assigns truth-values to worlds. Let w be this world. The proposition *Socrates was wise* assigns truth to w. Just as *x is a greatest teacher of Plato* is a concept of Socrates, not of the truth-value that it assigns to Socrates, so *Socrates was wise* is a concept of w, not of the truth-value that it assigns to w. Propositions are concepts of

worlds, not of truth-values.

When we make it clear what concepts are – functions from entities to truth-values – it is no longer tenable to say that the semantic contribution of a sentence-token is a truth-value.

Frege and Church would admit that "Socrates was wise" has different truth-values, and thus refers to different things, in different possible worlds. They didn't talk in terms of possible worlds, but that doesn't really affect the point. They would admit that *if* there were a world where English was spoken, but where Socrates was not wise, then "Socrates was wise" would be false. They would thus maintain that, in such a world, "Socrates was wise" would refer to the truth-value *false*.

Thus, Frege and Church would grant that "Socrates was wise", with its actual semantics, could have different truth-values, and thus different referents, in different worlds. They would admit that an expression with a fixed semantics can have different referents in different contexts. They could deal with this by saying that what a term refers to is not part of its semantic content. But that is absurd, as we have seen.

To avoid incoherence, they would be forced to go two-dimensional. The sentence-type is what keeps the same semantics, but that type doesn't refer to anything. *Tokens* of it are what refer. But once this is granted, it follows that sentence-tokens refer to their meanings. The semantic value of the type "Socrates is wise" is a function from contexts of evaluation to propositions. The semantic content of a token of "Socrates is wise" is some proposition. An *intension* is a function from contexts of evaluation to semantic values. An *extension* is the semantic value assigned to a token by the corresponding intension. Tokens are what have extension; types are what have intension. The intension of the type "Socrates is wise" is not a proposition, but a function from contexts of evaluation to propositions. The *extension* of a given

token of that type is a proposition. The extension of a sentence-token is *not* a truth-value.

Chapter 7 Reference as Semantic Contribution

If sentence-tokens of the form "…Plato…" did not bear propositions of the form:…*Plato.*…then "Plato" surely wouldn't refer to Plato. If sentence-tokens of that form *did* bear propositions of the form…*Charlemagne*…., then surely "Plato" *would* refer to Charlemagne. So, it would appear, "Plato" refers to Plato exactly if sentence-tokens of the form "…Plato…" bear propositions of the form:…*Plato.*…It thus seems to follow that "Plato" refers to Plato exactly if "Plato" *contributes* Plato to the meanings of sentence-tokens in which that expression occurs. We might put it thus: "Plato" refers to Plato exactly if "Plato" *semantically contributes* Plato. (x "semantically contributes" y exactly if y is what x contributes to the meanings of the sentences tokens in which x occurs.)

Of course, what we said about "Plato" and Plato generalizes without limit. For any expression E and any object O, E refers to O exactly if sentence-tokens of the form "…E…" bear propositions of the form:…*O.*…So E refers to O exactly if O is what E contributes to the meanings of the sentence-tokens in which E occurs. E refers to O exactly if E *semantically contributes* O. Thus reference seems indistinguishable from semantic contribution.

My view is exactly that – reference is indistinguishable from semantic contribution. If that is right, it immediately follows that token of "Plato

snores" or "I am tired" semantically contributes its *meaning*.

Sentence-tokens semantically contribute their meanings. This is a truism. So *if* reference is semantic contribution, then it immediately follows that sentence-tokens *refer to their meanings.*

But obviously some bridges have to be crossed if we are to make a case for the view reference collapses into meaning and that, consequently, sentence-tokens refer to their meanings.

Types versus Tokens

To start, we must once again distinguish expression-types from expression-tokens. The *type* "you" doesn't refer to anything. Tokens of it are what refer. Given this, we must say that the *type* "Plato" doesn't refer to anything, and that tokens of it are what refer. The content of the *type* "you" is a function from contexts to individuals. The content of a token of "you" is an individual. We must say the same thing about "Plato". The semantic content of the *type* "Plato" is a function from contexts of utterance to individuals. The semantic content of a token of that type is an individual.

Obviously there is a difference between the type "you" and the type "Plato". But the difference is *not* that the latter type refers while the former does not. The difference is that the semantic content of the latter is a *constant* function, while the semantic content of the former is a variable function.

We encountered an exact analogue of this problem before. The sentence-type "I am tired" doesn't encode a proposition – tokens of it are what do that.

It is pretty clear that "Plato" and "you" are in the same basic semantic (and grammatical) category. They are both singular terms. Suppose we said that the *type* "Plato" referred to Plato. In that case, the class of singular terms would comprise two radically different kinds of entities. If the type "Plato"

refers, then “Plato” is a very different sort of thing from “you”. The former would have an individual for its semantic content; the latter would have a function for its content. Under this circumstance, “Plato” and “you” would have altogether different semantic contents and, presumably, would therefore belong in altogether different semantic categories. But plainly they belong in the same basic semantic category. The differences between them are sub-categorical, not categorical.

The most natural approach is to say that the *type* “Plato”, no less than the type “you”, has a function for its semantic content. Obviously there is a difference between “Plato” and “you”. Tokens of “Plato” always refer to Plato. But there is no one person to whom all tokens of “you” refer. We accommodate this by saying that the semantic content of “you” is a variable function, while the content of “Plato” is a constant function. We *don’t* accommodate this by saying that the semantic content of the type “Plato” is an individual, while the semantic content of the type “you” is a function from contexts to individuals.

We encountered an exact analogue of this situation earlier. There is no denying that the type “I am tired” doesn’t encode a proposition. But other sentence-types *do* seem to encode propositions. The type “1+1=2” *does* seem to encode a proposition, and not a function from contexts of evaluation to propositions.

There are two ways we could deal with this. We could say: the type“1+1=2” has a proposition for its semantic content, while the type “I am tired” does not. Under this circumstance, those types would have entirely different kinds of thing for their semantic contents, and they would thus belong in entirely different semantic categories.

The natural approach is to say: The type “1+1=2”, no less than the type “I am tired”, has for it semantic content a function from contexts of evaluation

to propositions. The difference is that the one function is a constant function, while the other is a variable function. If we take this approach, we avoid taking the profoundly arbitrary view that some sentence-types encode propositions, while other sentence-types encodes functions from contexts of evaluation to propositions.

There is another, deeper reason to suppose that the type "Plato", no less than the type "you", has a function for its semantic content. It is obvious that in *some* cases, reference is a property of *tokens* - of spatiotemporal entities. This is not a pragmatic epiphenomenon: it is an integral part of the *semantics* of singular terms like "you" and "I" and, possibly, "the kind of France". So, uncontroversially, where some singular terms are concerned, reference is a property of *tokens*, not types. So, uncontroversially, reference is sometimes a feature of spatiotemporal entities – a property of (appropriately contextualized) deposits of ink and bursts of noise. Intuitively, it is very hard to believe that one and the same semantic function could be discharged *both* by spatiotemporal entities *and* by platonic entities. So it is very hard to see how the *type* "Plato" could refer. It is, to say the least, very unusual for a single task to belong both to the particular (the spatiotemporal) and to the universal (the platonic). So given that, in some cases, reference belongs to the particular, it would seem that it must *always* so belong. It thus seems *de rigueur* to assume that *tokens* of "Plato" are what refer, and that the corresponding type doesn't refer to anything.

So semantic content is *not* always reference. The *type* "you" and the *type* "Plato" obviously have semantic content, but they don't refer. When I say that "reference is indistinguishable from semantic contribution", I am referring to expression-*tokens*. So the principle I wish to defend is this:

(SC) Where expression-*tokens* are concerned, reference is identical with

semantic contribution – expression-tokens refer to their meanings.

Definite Descriptions in Relation to (SC)

Tokens of “the inventor of bifocals” presumably refer to Benjamin Franklin. In any case, this is the obvious thing to say about them. Russell famously denied this. He said that “the inventor of bifocals” (he didn’t distinguish tokens from types) does *not* refer to Benjamin Franklin. Russell had very good reasons for taking this view.

Right now I wish to show only that (SC) is entirely neutral on the question whether Russell is right. (SC) is consistent with the view that definite descriptions are singular terms, and it is also consistent with the view that they are quantifiers. In fact, Russell agrees with (SC): his agreement with it is the very thing that motivates the theory of descriptions.

I am not saying that “the inventor of bifocals” *does* in fact refer to Franklin. I am saying that *if* it refers to Franklin, that amounts to this. Sentences of the form “…the inventor of bifocals…” encode the proposition: …*Benjamin Franklin*…

So *if* “the inventor of bifocals” refers to Franklin, an immediate consequence is that

(i) “the inventor of bifocals was wise”

encodes the very same proposition as:

(ii) “Benjamin Franklin was wise”.

They both encode the proposition:

(iii) Franklin was wise.

Understandably, Russell thought it obvious that (i) and (ii) encode different propositions. He also rightly thought that *if* "the inventor of bifocals" referred to Franklin, then (i) and (ii) would both mean (iii).

To deal with this, Russell denied that "the inventor of bifocals" referred to Franklin; he denied that definite descriptions refer to individuals.

I think Russell's position is based on a confusion. Clearly (i) and (ii) *convey* different information. But it doesn't follow, and it probably isn't the case, that they have different propositions for their literal meanings.

But that is neither here nor there. Here what is relevant is that *if* "the inventor of bifocals" referred to Franklin, then Franklin, and Franklin alone, would be the semantic contribution of that expression.

Russell's theory of descriptions is not opposed to that. On the contrary, the theory of descriptions is a way of dealing with that fact.

Let us now deal with the main challenge to (SC):

Maybe you're right to say that tokens of "Plato" refer to Plato exactly if Plato is semantically contributed by such tokens. So in *some* cases meaning collapses into reference. But SC says that meaning *always* collapses into reference, and this is plainly false. Consider the word "and". This word has *meaning.* But surely it doesn't refer to anything. It has meaning by virtue of having some kind of *syntactic* of *formal* function. It does not have meaning by virtue of picking out some object.

You have a Neanderthal conception of meaning: to mean is to denote. This is exactly the position that Wittgenstein refuted in both the *Tractatus* and in

his later works. In the *Tractatus*, he refuted it in connection with connective terms like “and” and “or”. In his later work, he refuted it in connection with lexical items like “Nixon” and “Socrates”.

Semantic content is not *always* identical with reference. Expression-types, I have argued, do not refer to anything; but they have semantic contents: functions of some kind.

But with that qualification, I do have exactly the conception described. I thus hold that tokens of “and”, and morphemes like “-s” *do* refer. When we say that some expressions have a purely “formal” or “syntactical” function, we are really saying that they denote special kinds of things: second-order functions, to be precise.

Consider the word “and”. The standard view about it is this:

You don’t define “and” by pointing to some object that it denotes. You define it contextually. You say what it means by saying what is meant by whole sentences containing it.

The idea seems to be that if “and” referred to something, you could define it non-contextually; you could simply indicate what it denoted. Since it must be defined contextually, it doesn’t refer to anything.

This view is sheer folly. For “Smith” to pick out Smith is *precisely* for “Smith” to admit of a certain kind of contextual definition. It is precisely for “…Smith…” to mean:…*Smith*…

Of course, “Smith” can be defined ostensively, whereas “and” cannot. But ostensive definition is a form of contextual definition. When you say:

(*) *That* person is named “Smith”

that is just a condensed way of giving a contextual definition. You are saying:

(**) Consider *that* person over there. Let O be that entity. “…Smith…” means:…*O*…

This can be shown by analogues of arguments already given. Suppose you point to some object O and you say:

That person [pointing at O] is named “Smith”. But if you want to make a statement about him, you cannot use the word “Smith” to do so. “Smith smokes” doesn’t mean that *he* [pointing at O] smokes; it means that some other person smokes. In general “…Smith…” doesn’t mean that *he* [pointing at O] has:…x…

That would be absurd. You started out by giving an ostensive definition. But you then stripped that definition of any force it initially had.[270] So if you point to O and say “*that* is Smith”, a necessary condition for your definition to have any force is that “…Smith…” mean:….*O*…

That is also a sufficient condition. Suppose you said:

Sentences of the form “…Smith…” mean that *that* person [pointing to O] has:…x…But “Smith” does not name that person.

That would be absurd, for now familiar reasons.

When you give an ostensive definition, you are really giving a condensed contextual definition.

Of course, pointing cannot be involved in a definition of “and” or “or”. But that has nothing to do with semantics; that has to do with metaphysics and epistemology. We cannot point to abstract objects or to objects existing in the future. But we can refer to such things.

Perhaps we must define “and” contextually. But this is of no semantic consequence. “Smith”, no less than “and”, is always defined contextually.

Strictly speaking, we can no more point to the meaning of “Smith” than we can point to the meaning of “and”. When we say that Smith is the meaning of “Smith”, we are really making a statement about a class, an infinitely large class, of expressions. Smith’s being referred to by “Smith” – Smith’s being the meaning of “Smith” – is really identical with the fact that sentences of the form “…Smith...” mean:…*Smith.*…We can more point to *that* fact than we can point to the meaning of “and”.

The other side of the coin is that any contextual definition can be turned into a non-contextual or denotative definition. Frege rightly said that “some man” does not denote some man. He concluded that it must be defined contextually:

(i) “…some man…” is true exactly if: *for some x, x is a man and…x.…*

This is what people typically say about “and” and case-markers, and other so-called non-denoting expressions. This is taken to imply that such expressions do not denote anything. But it is easy to convert any contextual definition into a denotative definition. For example, (i) is equivalent to:

(ii) For any concept C, *C(some man)* is true exactly if for some x, x is a

man and C(x).

(ii) is equivalent to:

(iii) for any concept C, "some man" assigns truth to C exactly if for some x, x is a man and Cx.

(iii) is equivalent to:

(iv) "some man" denotes a function that assigns truth to a concept C exactly if, for some x, x is a man and Cx.

What we just said about "some man" can be done with *any* expression. Any contextual definition can be reduced to a denotative definition. If an expression can be defined contextually, it denotes something.

In effect, we've already seen this. Connectives like "after", "and", and "because" denote functions. The same thing is true of tense- and case-markers.

Granted, there is a difference between expressions like "Plato" and expressions like "and". But the difference is not that the former denote, while the latter do not. The difference lies in *what* they denote. The former denote individuals. The latter denote functions.

Actually, we will see some reason to believe that the difference is even *more* tenuous than this suggests. Arguably, tokens of "Plato" *do* denote functions, no less than "and". The difference between tokens of "Plato" and tokens of "and" lies in what kind of functions they denote.

The Augustinian conception of language, ridiculed by Wittgenstein, turns out to be quite right.

Since reference reduces to semantic contribution, it follows that sentence-tokens refer to what they semantically contribute; they refer to what they mean. In any case, if one is to escape this conclusion, one must show that SC is wrong. But this does not appear to be possible.

Some Loose Ends

I'd like to end this chapter by tying up some loose ends. Many (most famously Frege) have maintained that the meaning of a sentence is a function of the meanings of its parts. Linguistic meaning is "compositional". This is known as the "thesis of compositionality" or, simply, "compositionality".

Compositionality has been accepted by most, but not all. To my knowledge, it has always been regarded as substantive thesis. I think that compositionality is correct. But I also think it is a truism, in the same category as "we *can* travel in time: at the rate of one second per second".

We have argued that, for tokens of "Socrates" to refer Socrates just *is* for sentence-tokens of the form "…Socrates…" to have a certain meaning. If they *don't* have that meaning, then "Socrates" didn't refer to Socrates to begin with. There is no way that a sentence-token could not be a function of the meanings of its parts. An expression means such and such exactly if sentences containing it have certain meanings. So whatever meanings those sentences have, that fixes what the meaning of that expression is. Under no circumstance can we meaningfully, let alone truly, say that E means O but that "…E…" does not mean:…*O*…Compositionality is not a thesis *about* subsentential meaning; it merely identifies what subsentential meaning is.

Let us end this chapter by making a minor clarification. A token of "you", addressed to (say) Benjamin Franklin, semantically contributes Franklin himself, and not a concept that applies to him. This is also true of tokens of

"Benjamin Franklin" and possibly of "the inventor of bifocals".

The doctrine that some terms contribute *individuals*, as opposed to concepts applying thereto, is called "direct reference theory".

But really "direct reference theory" should just be called "reference theory". If tokens of "the inventor of bifocals" contribute the concept *inventor of bifocals*, as opposed to Franklin, then they would not refer to Franklin at all; they would be quantifiers, not expressions that referred to Franklin. Reference is always "direct".

The debate between friends and foes of "direct reference theory" is really a debate as to whether there is such a thing as reference; it is not a debate as to the nature of reference.

Chapter 8 The Causal Theory

So far, I've argued that:

(SC) E refers to O iff E semantically contributes O.

There is one challenge to (SC) that I have yet to consider. (SC) is in apparent contradiction with a very popular theory: the so-called "causal theory" of reference. That theory is:

(CTR) E refers to an object O iff tokens of E have a certain kind of causal relation to O (or states of affairs involving O).[271]

Here I will argue that (CTR) is *not* in fact opposed to (SC), at least not in so far as (CTR) is a coherent doctrine.

Some clarifications about CTR are appropriate. First of all, CTR is about expression-*tokens.* The expression-*type* "Socrates" isn't causally connected to anything. Tokens of it have causal relations. CTR says that such tokens refer to Socrates just in case they have a certain kind of causal relation to that man.

There are other, less innocuous, qualifications. No one would maintain that (CTR) is meant to apply to *all* referring terms. Consider "the first animal born in the year 3000 A.D." (If you think definite descriptions are quantifiers, then consider: "ze first animal born in the year 3000 A.D.")

Suppose that Alphie is the animal satisfying that description. In that case, Alphie is the thing referred to by that expression. But "the first animal born in the year 3000 A.D." refers to Alphie by virtue of encoding descriptive information that singles out Alphie; that expression does *not* refer to him by virtue of there being some causal relation between tokens of it and states of affairs involving Alphie. Such a causal connection doesn't exist. Even if it did, it wouldn't be what is operative.

This point can be generalized. (CTR) is not meant to apply to *any* expression that has constituent structure. "That snow is white" refers to the proposition *that snow is white.* The former refers to the latter because the latter is the thing having the properties in the descriptive content encoded in the former. CTR may be meant to apply to the constituents of "that snow is white", but not to that expression as a whole.

So far we've spoken about what CTR does *not* say. To explain what CTR *does* say, we need to define a few terms.

Let R be a referring term belonging to some language L. R is "rigid" iff it refers to the same thing in every world where L is spoken. In every world where English is spoken, "Richard Nixon" refers to Richard Nixon.[272]

Of course, there might be a world where "Richard Nixon" refers to John Kerry, but where *otherwise* English is spoken just as it is here. But, strictly speaking, the language spoken there isn't quite identical with the language spoken here; that version of English comprises the rule:

"Richard Nixon" refers to John Kerry.

Our version of English does not comprise that rule. So we are dealing with different sets of semantic rules, with different languages.

When we talk about rigid designation, we always have in mind worlds where the language in question is characterized by the *exact* same semantic rules that characterize it here.

There are two kinds of rigid designators. "Richard Nixon" refers to the same thing in all worlds where English is spoken. This is a matter of its *semantics.* The semantic rule for that expression is:

(RN) "Richard Nixon" refers to O exactly if O is identical with Richard Nixon.

"Richard Nixon" is what we call a *de jure* rigid designator.[273] Given

only its semantics, “Richard Nixon” must refer to Richard Nixon. Nothing non-semantic – not even facts about mathematics or logic – are needed to secure that connection.

“The square of two” is a rigid designator: it refers to the same thing in every world where English is spoken. But the *semantics* of that expression leave it open what it refers to in a different world. The semantic rule for that expression is:

(s) “the square of two” refers to a number n exactly if n is what results when two is squared.

The semantic rule is *not*:

(s*) “the square of two” refers to four.

In a world where English is spoken, but where the square of two is five, “the square of two” refers to five.

Granted, there probably is no such world. So “the square of two” refers to four in every world where English is spoken. But that is a fact about mathematics: the *semantics* involved leave it open whether “the square of two” refers to five or not.

“The square of two” is a *de facto* rigid designator. It does denote the same thing in all worlds where English is spoken. But this is a consequence of *non*-semantic facts; by themselves, the semantics of that expression do not guarantee that it will refer to four.

“Four” is a *de jure* rigid designator. The semantic rule for it is:

(FR) “Four” refers to O exactly if O is identical with the number four.

So no matter what laws of arithmetic obtain in a world w, “four” refers to four in that world, so long as English is spoken there; by themselves, the semantics of that expression guarantee that it refers to four.

(CTR) is meant, presumably, to apply to *de jure* rigid designators, like “Nixon” and “Socrates”. It is not meant to apply to *de facto* rigid designators like “the number of primes between 10 and 20”. *De facto* rigid designators refer by virtue of encoding the right descriptive information, never merely by virtue of a causal connection.

Even with all of these restrictions, problems arise for (CTR). That doctrine has a hard time dealing with expressions that are *de jure* rigid designators, such as “2” and “justice”. These terms denote abstract objects. Nothing has any causal connection to such an object. So it isn’t clear how CTR can deal with such terms.

I myself think that the causal theory is wrong even for “Nixon” and “Socrates”. But this is neither here nor there. What I wish to show now is that *even if* (CTR) is unrestrictedly true, there is still no opposition between it and (SC). There is no more opposition between those two views than there is between the view that a triangle is closed, three-sided, planar, straight-edged figure and the view that a triangle is the area bounded by three coplanar lines such that any two of them intersect, but not all three of them intersect.

The Right Causal Relation

The causal theorist must say *which* causal relations constitute reference. Not just any causal relation will do. If I push the door-bell, that causes the bell to ring. But the ringing of the bell doesn’t refer to my pushing of the button.

For the sake of argument, suppose that (CTR) is correct. Let R be the causal relation, whatsoever it should turn out to be, that constitutes or underwrites the relation of reference. In other words, R is the causal connection that causal theorists have in mind (or, better, that they are seeking).

How do we know whether a given type of causal connection is the right one? What is our benchmark? Suppose some not very intelligent causal theorist says:

(*) Consider the causal relation that holds between pushing the button at t and the bell's ringing at t*. Let us refer to that relation as DB. R is the kind of causal relation that holds between Socrates (or, more exactly, states of affairs involving him) and tokens of "Socrates" – it is the relation that makes those tokens refer to that person.

Everyone would *reject* (*). DB obviously does *not* do the job. If indeed tokens of "Socrates" stood in that relation to Socrates himself, that would *not* be enough for such tokens to refer to the man. Why not?

If you want to make a statement *about* some pushing of some button, you cannot use the resulting bell-ringing as a way of doing so. Suppose that the button is pushed at t, and the bell rings at t*. Let RING be that particular ringing, and let PUSH be the button-pushing in question. The sound-token:

(#) "[RING] [274] woke the dog."

does *not* bear the proposition:

(##) PUSH woke the dog.

(#) doesn't say *anything.* You have an unmeaning noise, followed by "is a distal cause of the dog's waking up". So (#) no more bears (##), or any other proposition, than does:

(###) "posidhfoasfhi woke the dog."

What RING is missing is the ability to play a role in sentences.[275] RING cannot be a part of a well-formed English sentence. That fact, by itself, bars us from saying that RING *refers* to anything, at least in English. In order to qualify as an expression of English that refers to PUSH, RING must be able to figure *grammatically* in English sentences. It would be absurd to say that a referring term *belonging to English* could not be used grammatically in English sentences.

An expression can be used grammatically in a sentence of English only if it can make a semantically significant contribution to such a sentence. So if it is to refer to RING, PUSH must be able to make such a contribution.

But that is not enough. The word "Morgan Freeman" makes semantically significant contributions to English sentences. But "Morgan Freeman" doesn't refer to PUSH.

If RING is to refer to PUSH, it is not enough that it make just any semantic contribution; it must contribute some specific thing. The thing it contributes must presumably be PUSH itself.

DB is the wrong causal relation because it does not enable RING to contribute PUSH to the meanings of sentences.

The other side of the coin is that, if DB did enable RING to contribute

PUSH to the meanings of sentences, then DB would be the right causal relation.

No causal theorist holds that any causal relation is reference-constituting. How does he decide which causal relation is the right one? The decision is based on considerations like the ones mentioned. Suppose x stands in causal relation R to y. If that is necessary and sufficient for x's semantically contributing y, then R is the right relation. If not, not. So R is the right causal relation exactly if it is consistent with SC.

A Qualification

RING could refer to PUSH. Indeed, RING could refer to PUSH in virtue of standing in a certain causal relation to PUSH. But that would not be enough to make the causal theory be correct.

Let L be a language that is just like English except that, by our stipulation, L is characterized by the following semantic rule. In L, the ringing produced by a door-bell pushing actually *refers* to that doorbell pushing. Let RING* be some particular door-bell ringing. In L, the spoken sentence-token "[RING*][276] woke the dog" means that a certain door-bell pushing woke the dog. [277]

Here we have a clear case where one thing refers to another in virtue of there being a causal connection between the two.

But notice that, by itself, the causal relation does not suffice for the reference relation. What suffices is that causal relation *plus* some semantic rule to the effect that if x and y stand in that causal relation, *then* sentences of the form "…x…" are to mean:…*y*…

The Content of the Causal Theory

This last point gives us some insight into what can legitimately be alleged by a causal theory of reference. This will, in turn, substantiate our view that reference is indistinguishable from semantic contribution.

For the sake of argument, suppose that the causal theorist is right; suppose there is some causal relation R such that x (a given sound or inscription) refers to y iff y stands in R to x.

In that case, tokens of "Socrates" refer to Socrates because Socrates stands in relation R to those tokens.

But it is easy to show a causal relation is never sufficient for reference. And when we make it clear what *is* sufficient, it turns out that a causal connection between object and token has a decidedly subordinate role.

Obviously there is an Albanian word for Socrates, a Finnish word for Socrates, an Urdu word for Socrates, and so on. These are different words. According to the causal theorist, for any token *t* of any one of these words, what makes *t* refer to Socrates is that Socrates stands in R to *t*.

The Urdu word for Socrates is "Sukrat". So according to the causal theorist, tokens of "Sukrat" refer to Socrates because Socrates stands in R to "Sukrat". The same is true *mutatis mutandis* of the Albanian word for Socrates, the Finnish word, the Hungarian word, and so on.

Let us focus on "Sukrat". If you were to put "Sukrat" in an English sentence (I mean, if you were to *use* N in an English sentence, not *mention* it), the result would be ungrammatical; you would not have a well-formed

sentence of English. If you were to say “Sukrat was wise” or “Sukrat drank hemlock”, the result would not be a well-formed sentence. People might understand you – after all, communication doesn’t always involve well-formed utterances. (If I child emits a certain kind of cry, I know that it wants food or wants its diaper to be changed. But its cry is not English or Urdu or any other language.) When you said “Sukrat was wise”, they might say: “You mean *Socrates* was wise! I didn’t understand you at first.” Their understanding you would involve their doing a bit of translation. This means that your utterance was not a well-formed utterance of *English*; it contained a renegade constituent, one that had to be replaced with an English word to yield proper English.

This point is liable to be misunderstood; maybe a different example would help. The Spanish word for “wet” is “mojado”. According to the causal theorist, tokens of “mojado” refers to the property of wetness exactly because they stand in a certain causal relation to instances of wetness. If you were to say “Smith is mojado”, what you were saying would be patently ungrammatical. The same thing is true of “Sukrat”, though perhaps less obviously. If you said “Sukrat was wise”, what you were saying would be ungrammatical. If you resist this, then (mutatis mutandis) simply replace “Sukrat” with “mojado” in the present discussion.

Of course, what is ungrammatical today might be grammatical tomorrow. Languages are plastic entities, at least along certain dimensions. Lexical items can be added quite freely. (Non-lexical changes – changes in syntax or grammar – are less easily made and are virtually never made intentionally.) So if you had a friend who always said “Sukrat” instead of Socrates, you would, in effect, *add* a rule to your idiolect to the effect that “Sukrat” refers to Socrates. But until such a rule were added to English – and, as of yet, it hasn’t been added – “Sukrat” is not in the English lexicon; it is not given a

meaning by the semantics of English; therefore “Sukrat was wise”, though perhaps easily enough understood, is not well-formed; it is not grammatical.

Given any referring term, it will be a referring term *of* this or that language. It will not be a referring term *simpliciter*. “Sukrat” refers to Socrates in Urdu, not in English. “Socrates” refers to Socrates in English, not in Urdu.

If the causal theory is right, then what makes tokens of “Sukrat” refer to Socrates is wholly that Socrates stands in relation R to those tokens. After all, the causal theorist says that x refers to y iff y stands in relation R to x. Tokens of “Sukrat” stand in relation R to Socrates. Therefore such tokens refer to Socrates.

But *in English* “Sukrat” obviously does *not* refer to Socrates. Quite possibly, in any language other than Urdu, “Sukrat” does not refer to Socrates.

But the causal theorist says that what makes tokens of “Socrates” refer to Socrates is simply that the latter stands in R to the former; the reference-relation in question is *constituted by* that causal relation. That causal relation is quite enough, according to the causal theorist, for “Socrates” to refer to Socrates in English. But the problem is that tokens of “Sukrat” stand in that relation to Socrates no less than tokens of “Socrates”. And yet tokens of “Sukrat” clearly *do not* refer to Socrates in English. So for x to refer to y in some language L, it is *not* enough that y stand in R to x. That may be a part of the story; but it cannot be the whole of it.

The causal theorist might resist this, and say:

The causal relation that an Urdu word must bear to Socrates, if that word is to refer to him, is different from the causal relation that an English word must bear to Socrates, if *that* word is to refer to him. Both of *those* relations are

different from the causal relation that a *Tagalog* word must be to Socrates, if that word is to refer to Socrates. Different causal relations are involved in different languages.

This position disjunctives the concept of reference. There is referenceinUrdu, referenceinEnglish, referenceinFrench, and so on.[278] But there is no longer a *single* concept of reference. This is plainly unacceptable. "Sukrat" refers to Socrates in the very sense in which "Socrates" refers to him, and in which the Chinese word for Socrates refers to him.

To avoid disjunctivizing the concept of reference, the causal theorist must say that some *one* relation R does the job. But we have already seen why such a position is untenable.

At this point, the causal theorist may weaken his position and say:

I never said that reference *is* some causal relation; I said that a certain causal relation is essentially *involved* in reference. So reference *per se* is not some causal relation; it is something else. That something else *involves* some kind of causal relation.

But then we are left with *no* answer to the question: "what is reference?" Yet this is the very question that CTR is supposed to answer.

A Place for CTR

Still, there is a certain truth in CTR. Consider the English semantic rule that is involved in the fact that "Socrates" refers to Socrates. That rule may well *exploit* some causal connection obtaining between certain current sound-

and inscription-tokenings, on the one hand, and states of affairs involving Socrates, on the other. That rule may be at least vaguely like:

(CRS) Let s be an arbitrary token of the physical type “Socrates”. If there is some object x such that x uniquely stands in causal relation R [for the appropriate value of R] with respect to s, then s refers to x. If there is no such object, then s refers to nothing. If there are many such objects, then s is ambiguous between them.

So given (CRS), and given that Socrates is the thing to which my token of “Socrates” stands in relation R, it follows that Socrates is what that token refers to. If Plato were the thing to which that tokening stood in that relation, then that tokening would refer to Plato.

In general, causal connections are involved in reference only when semantic rules exploit them.

This insight enables us to circumvent some problems associated with CTR. We can obviously refer to things to which we have no causal connection – for example, future individuals and abstract objects.

But there is only one concept of reference. “2” refers to the number two in exactly the sense in which “Nixon” refers to Nixon. There is no causal connection between the first pair. So CTR disjunctivizes the concept of reference.

Here we are dealing with a *different* disjunctivization of reference from the one discussed before. So CTR is guilty of doubly disjunctivizing an obviously unified concept. Thus CTR, if taken as a thesis as to what *constitutes* reference, is doubly wrong.

Advocates of CTR have tried to work around this problem, saying things like:

Tokens of “2” are causally connected to *instances* of the number two; and that is why such tokens refer to that number.[279]

But such maneuvers are obviously implausible. One can refer to the number two *itself*, not just to instances of the entity; and we have no causal connections to that entity itself.

Further, causal connections aren’t fine-grained enough to constitute-reference: I can refer to the clay, the statue, time-slices of the statue. I am referring to a different thing in each case. But any causal connection I have to the one is equally a causal connection to the others.[280]

Even if these problems can be finessed, CTR cannot possibly be true of referring terms that have phrasal structure, like “that Socrates wept often” and “the [ze] center of mass of the galaxy”. Such terms refer by virtue of describing, not by virtue of a casual connection. If CTR says that causal connections *constitute* reference, instead of being merely involved, then it says that the sense in which “Nixon” refers to Nixon is entirely different from the sense in which “the phi” (or “ze phi”) refers to the phi.

These problems vanish when CTR is put in its proper context.

In some cases, we define a word by *pointing* to its referent: “*that* man over there is Thaddeus Smith.” Reference is secured through ostension.

This is not always how words are defined. When we leave aside ostensive definition, we find that reference is always secured in the same way. We *describe* the thing to which an expression is meant to apply. But we give the description wide-scope. That way the sole semantic content of the expression is the individual: the expression doesn’t become a Russellian description-

quantifier.

Suppose I want to produce a term that refers to the first individual born in the 3rd millennium. I say:

(3M) *If* there is some object O such that O is the first individual born in the 3rd Millennium, then "Simon" refers to O.

Because the descriptive content in that rule is given wide-scope with respect to the "refers to" operator, the statement "There are possible worlds where Simon is born in the 2nd millennium" is true. This is what we want: after all, any individual *could* have had a slightly different birth-date.

As we've seen, in so far as CTR has any chance of being right, it is because certain semantic rules *exploit* causal connections. If a causal connection between tokens of "Socrates" and Socrates is involved in the fact that the former refer to the latter, it is because some rule like the following exists:

(S) If there is some individual O such that tokens of "Socrates" stand in R to O, then those tokens refer to O.

(S) and (3M) are completely parallel. In both cases, reference is secured through descriptive information. But that information is given wide-scope with respect to the "refers to" operator. Where (S) is concerned, the descriptive content has to do with causal relations; where (3M) is concerned, this is not so. But that doesn't have any special ramifications. Descriptions

differ; some involve the concept of causality; some don't. Some involve the concept of wearing a feathered-hat; some don't. I can describe an entity in terms of its causal liaisons or in terms of what kinds of clothes it is wearing. CTR merely registers the banal fact that, in some cases, the descriptive information in question has to do with a thing's causal liaisons.

Once this is realized, we don't have to invent tortured hypotheses as to why there really is a causal connection between the number two and tokens of the expression "2". There is no such connection; such tokens are not causally connected to the platonic type *two*, but only to its instances. Even if there were such a causal connection, it wouldn't be reference-constituting. What would be reference-constituting would be a semantic rule that fixed the referent of "2" in terms of descriptive information. That rule might exploit a description that alluded to the causal liaisons of certain things (although, in actuality, it would not). But that doesn't warrant a causal theory of reference. If I say "Jim is the man wearing the feathered hat", I am exploiting the fact that Jim is wearing a certain hat. In effect, I am saying: "there is some contextually salient object O that is wearing a feathered hat: 'Jim' refers to O." But this doesn't warrant a feathered-hat theory of reference.[281] [282]

A Nicety

Uncontroversially, it is *possible* that "Sukrat" refers to Socrates only in Urdu. We've seen how this fact causes trouble for the causal theory.

In my view, it is *analytic* that "Sukrat" can refer to Socrates only in Urdu. Given any word w of any language L, it is analytic that w cannot be used grammatically in a sentence not belonging to L. A homonym of w can be so used, but not w itself. Perhaps a homonym of "Sukrat" refers to Socrates in some language other than Urdu. In that case, that homonym can be used

grammatically in some language other than Urdu. But the *Urdu* word "Sukrat" cannot refer to Socrates, or have any grammatically licensed role, in any language other than Urdu.

Suppose the gases brewing in a volcano yield the noise "Socrates was wise". The volcano hasn't really said anything. It produced sounds that *would* have meant something if *we* had produced them. When we say "Socrates was wise", it is in consequence of our following certain semantic rules. If we were following no such rules, then we wouldn't be using a language at all; we wouldn't be saying anything. By exactly similar reasoning, what makes a sound be a token of a word is that one produces that token in consequence of one's following a certain semantic rule; if one is just *caused* to blurt it out – the way a corpse might be electrically stimulated to make noise – one wouldn't really have produced a bit of language at all. Unless we have something that is a consequence of one's awareness of a semantic rule, we don't have a word-token.

A corollary is that *which* word you are tokening is not determined solely by the phonetic properties of your utterance. If you really produced a word-token, and not just a noise that sounds like a word-token, that means you were following a semantic rule, and that you produced that word-token as a consequence thereof. Which word you tokened is a function of which rule you were following.

As we said before, a word never refers *simpliciter*; it always refers in this or that particular language. The semantic rule for "Socrates" is not: "Socrates" refers to Socrates. It is: "Socrates" refers to Socrates *in English.*

Suppose a speaker of L produced the sound "Sukrat" in consequence of following the semantic rule of *L* that assigns Socrates to "Sukrat". What he has tokened is a word of L, not of Urdu. He was not following an Urdu semantic rule, and thus didn't token an Urdu word. So strictly speaking, his

utterance was ungrammatical, even though that ungrammaticality is completely undetectable.

Chapter 9 Do Sentences Refer to Propositions?

Given only that the Slingshot fails, it doesn't follow that sentences do *not* refer to truth-values. It follows only that Frege and Church haven't shown that they do.

But in light of what we've seen, it is extremely unlikely that sentences refer to truth-values. The referent of an expression is the thing it contributes to sentences in which it occurs. Whenever "snow is white" occurs in a sentence, it contributes: *snow is white*. It doesn't contribute the truth-value *true.*

Consider any sentence of the form "…snow is white…" Examples are "grass is green and snow is white", "necessarily, snow is white", "John believes that snow is white", and simply "snow is white".

Surely all those sentences are about the proposition *snow is white.* Granted, those sentences are also about other things. But there can be no denying that, in virtue of containing the sentence "snow is white", a sentence is about the proposition *that snow is white*.

Reference is semantic contribution. Sentence-tokens semantically contribute their meanings. So sentence-tokens refer to their meanings. Their meanings are propositions. So sentence-tokens refer to propositions.

There is independent support for this. One cannot know what the semantic content of a referring term without knowing what it refers to. To grasp the semantic content of a referring term, one must know what it refers to. To grasp the semantic content of a sentence, or sentence-token, there is only one

thing one must know: what it means. One must know what proposition it encodes. So it is beginning to seem that sentences refer to their meanings. They refer to propositions, not truth-values.

Here is a disclaimer is in order. *Part* of what a sentence-token refers to is a proposition. But that is not the end of the story. There are important semantic differences between sentence-tokens that semantically contribute, and thus refer to, the very same proposition. Obviously "Plato snores" and "does Plato snore?" are semantically different, even though they both encode the proposition *that Plato snores.* A thoroughly correct semantics must account for this difference – and that is exactly what we will do. I will argue that "Plato snores" and "does Plato snore?" *do* refer to propositions: but that is not *all* they refer to. Each refers to a function *involving* a proposition. But they refer to different functions; and that, I will argue, is why they have different forces. I will argue that, notwithstanding years of tradition to the contrary, differences in force are differences in referent.

The best way to develop an appropriately nuanced semantics is to consider Frege's views on indirect discourse. This will clarify the relevant issues, and this will also give independent support to our view that *part of* sentence-tokens refer to are propositions.

Chapter 10 Frege on Indirect Discourse and Compositionality

Some contexts are *truth-functional.* Consider the sentence-token:

(1) "snow is white and grass is green"

Replacing "grass is green" with another sentence-token of the same truth-

value results in a sentence-token that has the same truth-value as (1):

(2) “snow is white and Frege defined numbers as sets of sets”.

A limiting case of such a context is a single sentence-token. If we replace

(3) “grass is green”

with a sentence-token of the same truth-value, what results has the same truth-value as the original. In (1) and (3), “grass is green” occurs truth-functionally.

Consider:

(4) “Little Timmy believes that grass is green”.

If we replace “grass is green” with a sentence-token of the same truth-value, what results doesn’t always have the same truth-value as the original:

(5) “Little Timmy believes that Frege defined numbers as sets of sets”.

In (4), “grass is green” does not occur truth-functionally.

Frege concluded from this that “grass is green” has one meaning when it occurs by itself, or in “grass is green and snow is white”, and a completely different meaning in “Little Timmy believes that grass is green.” In (3), “grass is green” semantically contributes a truth-value; in (4), it semantically contributes a proposition.

In general, for Frege, a sentence means one thing in truth-functional

contexts, and a different thing non-truth-functional contexts.

Frege had another reason for holding this. The meaning of a sentence is a function of the meanings of its parts. This is known as the principle of compositionality. (We discussed this principle earlier.) Frege firmly believed in this principle; and, unquestionably, it is correct.

Frege held that, in (3), the semantic contribution of "grass is green" is a truth-value. He recognized that, in (4), its semantic contribution is *not* a truth-value. If Frege said that "grass is green" contributed a truth-value in (4), then he would be forced to say that the meaning of (4) was to no degree a function of what "grass is green" meant in that sentence. In other words, Frege would be forced to deny compositionality.

But Frege rightly wanted to hold onto compositionality; he wanted to say that the meaning of (4) was a function of the meanings of its parts. So he was forced to say "grass is green" means one thing in (3) and a different thing in (4).

Intuitively, Frege's analysis seems very wrong. Surely "grass is green" has the very same meaning in all of:

(3) "Grass is green"
(4) "Little Timmy believes that grass is green."
(6) "It possible that grass is green".

In all these cases, "grass is green" contributes the proposition *that grass is green.* Any correct semantics must be consistent with this fact. If Frege's analysis is right, then in (3) "grass is green" contributes a truth-value, whereas in (4) and (6) it contributes a proposition. This strikes me as deeply wrong – as a position that must be avoid at all costs. It is pretty clear that "grass is green" has the *same* meaning in "grass is green and snow is white"

and also in “possibly, snow is white”. Frege’s position strikes me as one that somebody would hold only if he were *forced* to hold it.

There is additional reason to be suspicious of Frege’s analysis. It makes “grass is green” be *infinitely* ambiguous. It means one thing (3), a different thing in (4), yet a different thing in

(7) “Smith believes that Little Timmy believes that grass is green.”

Every addition of an epistemic (or otherwise non-truth-functional) operator generates a new meaning.

But surely “grass is green” is not ambiguous at all, let alone infinitely ambiguous. A correct semantics must do justice to this fact.

The Misconceptions Underlying Frege’s View

Two misconceptions underlie Frege’s astonishing position that “grass is green” means one thing in “grass is green and snow is white” and a completely different thing in “possibly, grass is green”.

We are familiar with one of these misconceptions. He argued that (in non-oblique contexts) sentences refer to truth-values. It is obvious that they *don’t* refer to truth-values in oblique contexts. Therefore, Frege was forced to say that “grass is green” is ambiguous – infinitely ambiguous in fact.

We have seen that the Slingshot doesn’t prove anything; in particular, it doesn’t prove that sentences refer to truth-values. So that eliminates one of Frege’s arguments.

Another misconception underlies Frege’s strange view. Frege is, of course, right to say that some contexts are truth-functional while others are not. But he is wrong to say that for a sentence to occur truth-functionally is for it to

contribute a truth-value. That idea is based on a massive non-sequitur. Whether “Kermit is green” is a *function* of whether Kermit has the property of being green. But “Kermit” obviously doesn’t *contribute* that property. “Kermit” contributes the thing to which that property is *ascribed*: it doesn’t contribute the property itself. Whether “John believes that grass is green” is a *function* of whether the proposition *grass is green* has the property of being believed by John. But obviously “that grass is green” doesn’t *contribute* that property. It contributes the thing to which that property is being ascribed: it doesn’t contribute the property itself.

Whether “snow is white and grass is green” is true is a function of whether the proposition contributed by “grass is green” is true. But this in no way implies that “grass is green” itself contributes that truth-value. It implies that “grass is green” contributes the thing *to which* that truth-value is being ascribed.

What is implied is actually that “grass is green” does *not* contribute a truth-value. To say that “Kermit is green” is a *function* of whether the thing denoted by “Kermit” is green is to say that “Kermit is green” ascribes the property of being green *to* that thing. So the thing denoted by “Kermit” must be *distinct* from the aforementioned property. To say that the truth of “grass is green and snow is white” is a function of whether the thing denoted by “grass is green” is true is to say that it ascribes the property of being true *to* that thing. So the thing denoted by “snow is white” must be *distinct* from the aforementioned property. So the fact that “grass is green” occurs truth-functionally in that sentence is to say, in fact, that it does *not* contribute a truth-value: it contributes the thing *to* which a truth-value is ascribed, and thus does not contribute the truth-value itself. Thus Frege’s is simply spurious. Given that “grass is green” occurs truth-functionally in some sentence, it does *not* follow that it contributes a truth-value. Actually, it

follows that it does *not* contribute a truth-value: it contributes the thing *to which* a truth-value is ascribed.[283]

Indirect Discourse: a Positive Account

Frege *did* make a point about sentence-reference that is, I think, quite correct and that forms the basis of a correct analysis of indirect discourse. He said that, in

(4) Little Timmy believes that grass is green,

"grass is green" (or, strictly, "that grass is green") refers to the proposition: *that grass is green.*

A correct account of indirect discourse merely generalizes this point of Frege's. It extends it to contexts where, if we over-value phonetics, it seems not to apply.

Let us consider our paradigms:

(i) "Plato snores".
(ii) "Fred believes that Plato snores."
(iii) "It is possible that Plato snores."
(iv) "that Plato snores"

Let (i)-(iv) be expression-tokens, not expression–types. Obviously each of (i)-(iv) semantically contributes the proposition *that Plato snores.* This suggests that each comprises a constituent that refers to that proposition. Of course, it is immediately obvious that each of (ii)-(iv) comprises such a constituent. The problem-case is (i). But (i) is not a problem case, provided that we don't put too much semantic stock in phonetic surface structure. I would suggest that, semantically if not phonetically, (i) *does* comprise an occurrence of "that Plato snores". I would also suggest that, semantically if not phonetically, (i) comprises a constituent that ascribes *truth* to the proposition denoted by "that Plato snores".

So here is the picture I am presenting. Semantically, if not phonetically, each of (i) –(iv) comprises an occurrence of "that Plato snores". That expression denotes the same thing in all of (i)-(iv): the proposition *that Plato snores.* What varies from case is to case is what is being said *about* that thing. Nothing undergoes a shift in meaning.

Let us start with the obvious.

(ii) comprises an expression ("that Plato snores") that refers to the proposition *that Plato snores*. (ii) also comprises material that ascribes the property of being believed by Fred to that proposition. Consequently, (ii) is true just in case that proposition has that property.

(iii) comprises an expression ("that Plato snores") that refers to the proposition *that Plato snores*. (iii) also comprises material that ascribes the property of being possibly true to that proposition. Consequently, (iii) is true just in case that proposition has that property.

(iv) comprises an expression ("that Plato snores") that refers to the

proposition *that Plato snores*. (iv) doesn't comprise any material other than that expression; it doesn't ascribe any properties to that proposition or to any other proposition. Consequently, (iv) isn't true or false.

Given these obvious facts, here is what I would say about (i).

(i) comprises an expression ("that Plato snores") that refers to the proposition *that Plato snores*. (i) also comprises (phonetically unrealized) material that ascribes truth to that proposition. Consequently, (i) is true just in case "that Plato snores" denotes a true proposition.

The only reason to question this analysis is that *at the level of phonetics*, (i) doesn't comprise the expression "that Plato snores" in its entirety. But phonetics is an extremely untrustworthy guide to semantics. Sentences are replete with phonetically unrealized constituents, and semantics would be impossible to do without granting their existence.

Here is an extreme example. You ask me "who was the greatest philosopher of antiquity?" In response I say "Plato". Obviously my response is elliptical for: "Plato was the greatest philosopher of antiquity." But, for the most part, the semantic content of my response was suppressed. If the last fifty years of linguistics are to be given any credence, this kind of ellipsis pervades language.

Apart from phonetic surface-structure, there is no reason to suppose that (i) doesn't comprise an occurrence of "that grass is green" and that it doesn't comprise material ascribing truth to the corresponding proposition. And there is every reason to suppose that it *does* comprise such material. If we make

this supposition, we have an immediate explanation of all the relevant facts. We have an immediate explanation of why all of (i)-(iv) semantically contribute the proposition *that Plato snores.* We also immediately have an explanation of why the truth of (i) is dependent on the truth of that proposition, whereas this is not the case with (ii) and (iii).

If this is right, then there is no constituent of (i) that has one meaning in (i) and a different one in (ii) or (iii) or (iv). Nothing has shifted meaning. The only shift relates to what has been *said* about the thing denoted by "that Plato snores".

"Plato snores" occurs truth-functionally in (i), and non-truth-functionally in (ii), because in (i) truth has been ascribed to the corresponding proposition, whereas in (ii) truth has *not* been ascribed to it.

Here is another way to think about it. Consider the sentence-tokens:

(i_T) "*That Plato snores* is true."
(ii_T) "*That Plato snores* is believed by Fred.
(iii_T) "*That Plato snores* is possible."

Obviously "that Plato snores" means the same thing in all of these cases. Also, it is clear that (ii_T) is a mere re-wording of (ii), and that (iii_T) is a mere re-wording of (iii). Obviously "that Plato snores" occurs truth-functionally in (i_T). Clearly, it does not mean something different from what it means in (ii_T) and (iii_T). The fact that the one context is truth-functional, while the other two are not, doesn't have anything to do with a meaning-shift. So there is no *general* inference from "S is occurring non-truth-functionally" to "S has undergone a shift in meaning."

Obviously (i_T) is very close to (i). It seems more natural to say that (i) is

a compressed form of (i) than it is to say that the words “Plato snores” have entirely different meanings. (This is not *quite* accurate: see below.)

Also, it isn’t good methodology to posit meaning shifts every time a context is non-truth-functional. “Kennedy died” occurs non-truth-functionally in “Kennedy died because he was shot”, “Kennedy died after Lincoln died”, “it is regrettable that Kennedy died”, and so on. Most connectives are non truth-functional. The only ones that are truth-functional are “not”, “and”, “or”, and “it is true that”. (I am leaving aside manufactured ones like “it is true that it is true that”, and “it is not the case that it is not the case that”.) There is no need to posit meaning-shifts, as long as we do not put too much stock in phonetics.

Am I Over-using the Concept of Deep-structure?

I should deal with a misgiving some may have about my analysis:

You talk very freely about “phonetically realized” sentence-constituents and, by implication, about “surface-structure” and “deep structure”. But it is quite possible that all such talk is hogwash. Granted, Chomsky and others talk and think this way. But it is very much an open question whether those people are right. Your commitment to Chomsky’s dubious paradigm weakens your analysis.

I leave it open whether Chomsky is right. But I don’t think my analysis

requires that he be right. My analysis has its roots in the very *concept* of a linguistic expression; its truth isn't contingent on the fortunes of an empirical doctrine like Chomsky's.

The English word "Socrates" can be tokened by physical entities that have absolutely nothing in common. First of all, that word can be spoken or written or expressed by hand-signals. No hand-signal has *any* physical similarity to any spoken token or any written token. No written token has any similarity to any spoken token.

Further, given any two written tokens of "Socrates", they may bear minimal resemblance to each other. If a court-reporter tokens that word in his short-hand, the result will bear little, if any, resemblance to what I write if I type it or if I write it in cursive.

If a person with damaged vocal chords tokens that word, the result bears no resemblance to what results when I token it.

What makes a physical entity count as a token of a certain word is not merely its physical properties – is not its shape, or its acoustical properties. Such properties have a role, but not the role one might think.

A corollary is that if an inscription perfectly isomorphic with the following

SOCRATES

were tokened, but it weren't tokened in the right cultural context, or in consequence of the right intentions, then it wouldn't be a token of the word "Socrates".

Suppose a coffee-machine were by sheer accident to produce the sounds "Socrates was wise". Would the machine have *spoken*? Would it even have produced any tokens of any words? No. It would have produced sounds that,

had *we* produced them, would have qualified as speech. The machine hasn't spoken at all; it hasn't said anything. It made sounds, but it was not using those sounds to execute some linguistic intention. So it didn't produce word-tokens. It produced sounds that, had we produced them, would have been words. Where the right psychological background is lacking, there are no word-tokens.

But the right psychological background is not enough. First of all, one never just speaks or writes; one always speaks or writes in this or that language. Now suppose I *think* that, in English, "blrak blurg gralb" means *grass is green*; and I say "blrak blur gralb" with the intention of saying *grass is green.* Because there is no existing practice of taking "grass is green" to mean *grass is green*, I haven't said anything, at least not in English. And I haven't said anything in any other language. As we saw earlier, when discussing "Sukrat", if a noise that I produce is to qualify as a token of a word of some language L, then that noise must be a consequence of my believing that *in L* that noise has a certain meaning. My utterance of "blrak blur gralb" was a consequence of my belief that *in English* that noise means *grass is green*; it was not a consequence of any beliefs about L. So I didn't say *anything* in any language.

What makes a physical event be a tokening of this or that expression is not that it has these or those phonetic properties. It is that the event is caused by certain psychological events, in a certain cultural or historical context.

The phonetic or geometrical properties of that event help others to *identify* the presence of the psychological and cultural factors that allow for the tokening of expressions. That seems to be their primary function. Up to a point, those phonetic properties may even be more than just indicators of such factors; up to a point, phonetics may be constitutively involved. But the presence of those phonetic properties is not, at least not by itself, what *makes*

that event be an actual expression-tokening, as opposed to something that just sounds or looks like one. Phonetics has a role; but it is not as central as might be thought.

Any expression can be physically realized in different ways. An expression-type like “that grass is green” is no exception. For the type “that grass is green” to be tokened is, at least in part, for something with a certain *semantics* to be tokened. What gives a physical event the right semantics is, at least in part, that it has the right psychological and cultural background. If that background is present, there is enormous latitude as to what the actual physical properties – the phonetics, the morphology – of that token can be. My theory asks that “that grass is green” have different phonetic (and acoustic) incarnations. In light of what was just said, this is not much to ask. [284]

Chapter 11 Refining our Position

For reasons just discussed, the meaning of a sentence-token is *in part* a proposition. So part of what a sentence-token refers to is a proposition. But that is not *all* it refers to. A correct semantics must say *exactly* what a sentence-token refers to.

Two sentence-tokens can encode exactly the same proposition but differ profoundly in their literal meanings.

(i) “Plato snores”

is true or false.

(ii) “Does Plato snore?”

is neither true nor false.

But both encode exactly the same proposition: *that Plato snores.*

Further, tokens of

(iii) “Plato, snore!”

and

(iv) “that Plato snores”

encode that very same proposition. But no two of them have quite the same literal meaning. (iv) isn’t a sentence-token at all. It doesn’t assert, command, or ask anything. But (iii) obviously does order command something.

A correct semantics must account for these differences among (i)-(iv).

We’ve identified reference with semantic contribution. What an expression-token refers to is nothing other than what it semantically contributes to sentences in which it occurs.

Tokens of (i) are assertions; they have indicative *force*. Tokens of (ii) are commands; they have imperative force. Tokens of (iii) are questions; they have interrogative force. Tokens of (iv) *that Plato snores* don’t have any of these properties.

So tokens of (i)-(iv) all contribute *different* things. If our analysis of reference is correct, they cannot all refer to the same thing.

At the same time, there is not the slightest doubt that at least *part of* what is semantically contributed by tokens of (i)-(iv) is the proposition: *that Plato smokes.* So *part of* what they refer to is that proposition.

The key to sorting all this out lies in a correct understanding of the concept of *force.*

(i) “Plato snores”
(ii) “Does Plato snore?”
(iii) “Plato, snore!”

(i)-(iii) obviously have different forces; and this is a profound *semantic* fact. For some reason, philosophers of language have all but totally neglected the concept of force. What has been said is either trivial or circular (“force is what is added by a force-operator”) or utterly implausible (“differences in force are pragmatic, not semantic”, “force operators don’t have any effect on the meaning of the sentences in which they occur”, “ultimately, all sentences have indicative force, for example (iii) is synonymous with: the following utterance of mine is to be taken as a command: you, Plato, will snore”). It is time to take the concept of force seriously. This, and only this, will make it possible to provide an adequate analysis of the concept of a sentence.

By way of anticipation, here is what I propose. For reasons we’ve already considered, part of what tokens of (i)-(iv) refer to is the proposition: *that Plato snores.* The reason is that all those tokens actually *contain* the expression: “that Plato snores”. (i)-(iii) contain it in a phonetically disguised form. Suppose somebody asks you “who was the greatest philosopher of antiquity?”, and you say “Socrates”. Your answer contains a phonetically disguised: “was the greatest philosopher of antiquity.” Similarly, (i)-(iii) contain a phonetically distorted “that Plato snores”.

But (i)-(iii) do not *just* contain that expression. They contain extra semantic material. In each case, that material denotes a function. That

function is what gives indicative force to tokens of (i), imperative force to tokens of (ii), and so on.

The extra material in (i) is responsible for the fact that (i), unlike (ii) and (iii), is true or false. It is responsible for the fact that (i) has assertoric force. So *very roughly speaking* that extra material might be thought of as ascribing truth to the proposition in question. (We will see why this is a very rough way of putting it.)

The extra material in (ii) is responsible for the fact that (ii), unlike (i) and (iii), is a question and is thus supposed to be given answer.

The extra material in (iii) is responsible for the fact that (iii), unlike (i) and (ii), is a command.

We will work out the details shortly.

So a correct semantics must account for the semantic differences among (i)-(iii). At same time, it must also do justice to the fact that "Plato snores", "does Plato snore?", and "Plato, snore!" are all members of the same basic semantic category. They are all sentences. There are important differences between them. This puts them in different *sub*-categories. But there is some one non-disjunctive category that comprises all and only those things we call "sentences". A correct semantics must find the property had by all and only the things in that category.

A sentence-token, I will argue, is an expression that denotes a very special kind of function.

I want to forestall a couple of mistaken views that might be had in connection with the matters just described. Then I will give my own analysis.

Force versus Content

How are we to explain the fact that

(i) “Plato snores”

is true or false, while

(ii) “that Plato snores”

is neither? How are we to explain the fact that they have *different* forces?

There is one obvious answer. One might say that they actually *do* contribute different propositions, and that this is why they differ in force. [285] One might say that, in general, differences in force *do* reduce to differences in propositional content. The idea would be this:

(i) contributes the proposition: *that it is true that Plato snores*, while (ii) merely contributes the proposition: *that Plato snores.* This is why (ii) is neither true nor false, while (i) is true or false. A difference in force *does* reduce to difference in the proposition contributed.

But that is absurd. Even if we grant that “Plato snores” contributes the proposition *that it is true that Plato snores*, as opposed to (merely) *that Plato snores*, that has nothing to do with the difference in *force* between “Plato snores” and “that Plato snores”. For the expression “that it is true that Plato snores” is neither true nor false. But, uncontroversially, the semantic contribution of that expression is the proposition: *that it is true that Plato snores.* Differences in *force* are not reducible to differences in propositional content.

Analogous remarks apply to

(iii) “does Plato snore?”

and

(iv) “Plato, snore!”.

What we just said (mutatis mutandis) shows that these sentence-tokens do *not* differ from “that Plato snores” in propositional content; they differ in some other, completely different respect.[286]

Force not merely a way of Signaling a Propositional Attitude

Here is a view that has elements of truth, but is wrong overall:

A force-operator is a conventionalized way of *expressing* an attitude towards a proposition. “Snow is white” encodes the proposition *that snow is white*. At some level, “snow is white” contains a force-operator. The presence of that force-operator is the conventional way of expressing one’s belief in that proposition.

An exactly similar line can be taken for questions and imperatives. “Is snow white?” can be seen as comprising a force-operator: one giving interrogative force. The presence of that operator is a conventional way of expressing one’s wish to know whether a certain proposition is true. The proposition, of course, is: *that snow is white.*

There may be truth in this position; but, as it stands, it is insufficient. Force-

operators probably *are* ways of indicating propositional attitudes. But that is not all they are. Consider the sentence:

(v) “I wish to know whether Plato snores.”

Surely that sentence is a conventionalized way of expressing a desire to know whether the proposition *that Plato snores* is true. But an utterance of (v) is semantically very different from an utterance of

(iii) “does Plato snore?”

Semantically, (v) either true or false. Semantically, (iii) is neither. So what makes (iii) be a question is not merely that it is a conventionalized way of expressing one’s desire to know whether a certain proposition is true.[287] That is probably part of the story, but not all of it.

Another example might help. Consider:

(vi) “I believe that Plato snores”

A token of (vi) is a conventionalized way of expressing one’s belief in the proposition *that Plato snores*. But a token of (iv) is semantically very different from a token of:

(i) “Plato snores”.

A token of (vi) encodes a proposition that is true just in case a certain individual has a certain belief. A token of (i) encodes a proposition that has

nothing to do with what anyone believes. (i) could be true in a world where nobody had any wishes. (vi) could not.

So it is not enough to say that force-operators are conventionalized ways of expressing propositional-attitudes.

Force a Matter of Literal Meaning

Another view I want to head off is this:

You say that there are *semantic* differences among tokens of “Plato snores”, “that Plato snores” “does Plato snore?”, and so on. I think this is false. Obviously they are all associated with the same proposition and the same truth-value. Obviously they all contribute the same proposition. The meaning of a sentence-token obviously just *is* the proposition it encodes. This is a truism. The meaning of a sentence-token just *is* its semantic content. This is a truism. So it follows from truisms that there are not *semantic* differences among “Plato snores”, “does Plato snore?”, and so on. Perhaps there are pragmatic differences: differences in what they *do* (perlocutionary differences). But there are no differences in what they *mean*. You’ve made a case that sentence-tokens semantically contribute propositions, and that reference reduces to semantic contribution. So instead of fiddling about with irrelevant pragmatic notions like sentence-force, you should just end your story right here.[288]

The problem with this proposal is that “Plato snores” quite plainly *does* differ in literal meaning from “does Plato snore?” and “that Plato snores”. Suppose you want to assert that Plato snores. If you say “does Plato snore?”

or “Plato, snore!”, you are obviously guilty of linguistic incompetence: you obviously don’t know quite what is meant by the expressions you are using. So we are not dealing with mere pragmatic or perlocutionary differences. [289]

Here we have to distinguish two issues that have traditionally been conflated. At the level of literal meaning, a token of “Plato snored” is about Plato. But this doesn’t mean that when one *uses* such a token one is necessarily making a statement about Plato. Given the right circumstances, any expression, no matter what its literal meaning, can be used to say anything. Sentence-tokens are often used to convey the exact opposite of what they literally mean. (This is what sarcasm is about.) What a *person* says is not always what his *words* literally mean.

This obvious point has been used as the basis for a very wrong conception of sentence-force. The study of the making of assertions belongs to pragmatics. The literal meaning of a sentence-token obviously places *some* constraints on what people *use* that sentence to assert. But what is asserted is a function only in part of literal meaning. A theory of *assertions*, i.e. of the making of assertions, must obviously go beyond literal meaning.

What we just said about assertions applies to all other speech acts. At the level of literal meaning, a token of “does Plato snore?” is a question. But such a token can be *used* to do things other than ask a question. Given that right circumstances, one could use such a token to make a statement, issue a command, expression a wish, and so on. The study of the asking of questions must obviously go beyond literal meaning. Exactly similar remarks apply to the issuing of orders.

So the *making* of assertions is a topic for pragmatics; so is the asking of questions and issuing of commands. But assertoric *force* is a topic for semantics; the same is true of interrogative and imperative force. A

comparison might help. As we just saw, the *making* of statements about Plato is a topic for pragmatics. At the same time, it is obviously a matter of *semantics* that tokens of “Plato snores” are about Plato. Any theory that fails to register a semantic difference between “Plato snores” and “Socrates snores” is obviously a failure. It is irrelevant that the *making* of statements about Socrates and Plato is a topic for pragmatics. There is obviously a semantic difference between “Plato snores” and “does Plato snore?” and “Plato, snore!”. Any theory that fails to register this difference is obviously false. It is irrelevant that the making of assertions and the asking of questions and the issuing of commands belongs to pragmatics.

Semanticists have held the concept of force in contempt. But force-operators are exactly like other expressions. Tokens of such operators refer, no less than do tokens of “Plato” and “you”. [290] They refer to functions of a certain kind. But they are not the functions one might think. Once we become clear on the semantics of force-operators, we obtain real insight into problems that seem completely unrelated – in particular, problems relating to the semantics of expressions occurring in “oblique” contexts. In the next chapter, we will say exactly what it is that force-operators do.

Chapter 12 Force and the Semantics of Sentence-tokens

By a “speech act” I mean the tokening (the uttering or writing) of a sentence. If you say “did Plato snore?” or “Plato snores”, you are performing a speech-act.

Many different things can be achieved through speech acts. Money can be made, foes can be wounded, lovers can be wooed. Speech-acts bring us many kinds of success.

But very few such successes fall within the province of semantics. Semantics is concerned only with literal meaning. Suppose I say

(i) "Plato snores".

I can use these words to accomplish many different things - to amuse, threaten, or deceive. But none of this has any relevance to the *semantics* of my words.

But there is *a* kind of success associated with such an utterance that does fall within the province of semantics. It is not part of the *semantics* of an utterance of "Plato snores" that it must amuse or deceive. But it *is* part of the semantics of an utterance of (vi) that it is supposed to be *true.* It *is* part of the semantics of such an utterance that it is a failure if false, and a success if true.

Everyone who speaks English knows as much. Anyone who speaks English knows that an utterance of "Plato snores" is supposed to be true and, thus, that it is a success only if it is true.

In general, there is exactly one kind of success that, at the level of semantics, indicative sentence-tokens are *supposed* to have. They are supposed to be *true.*

This doesn't mean that people always aspire to tell the truth when they use indicative sentences. But it does mean that *semantically* they are failing if the indicative sentence-tokens they produce are false.

Under what circumstances is an utterance of "Plato snores" true? It is true exactly if a certain proposition is true. That proposition is: *that Plato snores*. Of course, *that Plato snores* is exactly the proposition that is semantically encoded in such utterances.

Let us take stock. At the level of semantics, there is exactly one kind of success that an utterance of "Plato snores" is supposed to have: it is supposed

to be true. Such an utterance is true exactly if the proposition semantically encoded in it (*that Plato snores*) is true. So the semantics of an utterance of "Plato snores" assigns it the property of success exactly if the proposition encoded in that utterance is true. Thus, the *semantic content* of such utterances assigns success to such utterances exactly if the propositions encoded in them are true.

Therefore the semantic content of an utterance "Plato snores" is a function that assigns the property of success to that very utterance exactly if the proposition *that Plato snores* is true.

This is easily generalized. Let *t* be any indicative sentence token. *Part* of what *t* will encode is some proposition P. *t*'s semantic content is a function that assigns the property of success to *t* exactly if P is true.

Force-operators

Utterances of "that Plato snores" are neither true nor false. This is because they do not have assertoric force. In fact, they don't have any kind of force.

Presumably such utterances lack any constituent that give them sentential force. They do not comprise force-operators.

Utterances of (i) "Plato snores" are true or false because they *do* have assertoric force. So they must comprise some constituent that gives them such force. They must comprise force-operators.

Force-operators cannot be identified with any isolable part of the phonetics of such utterances. This may make people leery of supposing that they comprise such operators. But this is extremely shallow reasoning. Sentences are replete with constituents that have no phonetic representation or a highly distorted phonetic representation.

So let us operate on the reasonable assumption that, semantically if not

phonetically, utterances of “Plato snores” *do* comprise force-operators.

What exactly is it that is given assertoric force?

(a) “Plato snores”.
(b) “Does Plato snore?”
(c) “Plato, snore!”
(d) “that Plato snores”.

Uncontroversially, *part* of what is semantically encoded in any token of (a)-(c) is the proposition *that Plato snores.*

Given that (a) semantically encodes (among other things) *that Plato snores*, it is not unreasonable to suppose that, semantically if not phonetically, it contains an expression referring to that proposition.

The same is true of (b) and (c). Semantically, if not phonetically, they comprise an expression referring to: *that Plato snores.*

Propositions are not given force. The proposition *that Plato snores* cannot be given interrogative, assertoric, or any other force. Expressions are given force.

So presumably what is going is this. In (a), an expression referring to the proposition *that Plato snores* is given assertoric force. In (b), such an expression given interrogative force. In (c) such an expression is given imperative force.

Further, given that (a) assigns assertoric *force* to the expression referring to the proposition just mentioned, it is not unreasonable to suppose that, semantically if not phonetically, it contains a force-operator.

So the right way to look at (a) is as a phonetic distortion of:

(a_F) “A(that Plato snores)”.

where the “A” is an operator that gives assertoric force to the proposition falling its scope.

Right now, let us focus on (a) and (a_F). We will return to (b)-(d) in a moment.

Any token of (a_F) includes (inter alia) a token of “that Plato snores”. That is why any token of (a_F) semantically encodes: *that Plato snores.*

But any such token also includes a force-operator. Obviously that force-operator gives assertoric force to the occurrence of “that Plato snores” in its scope. But what exactly does this mean? What is it to give assertoric force to an expression?

We’ve already considered some erroneous answers to this question. Let us now give a correct answer.

(a_F) “A(that Plato snores)”.

Let *t* be an arbitrary token of (a_F). Of course, *t* is (inter alia) a token of “that Plato smokes”. Any token of “that Plato snores” denotes the proposition *that Plato snores.* The “A” in a given token of (a_F) denotes a function that ascribes success to *t* exactly if the proposition just mentioned is true.

We must distinguish the sentence-*type* (a_F) from tokens of that type. The semantic content of that *type* is a function that, in its turn, assigns a certain kind of function to each token of that type. The semantic content of the type (a_F) is a function that assigns another function to any token *t* of that type; and that second function, in its turn, assigns the property of success to *t* exactly if a certain proposition (*that Plato snores*) is true.

Unquestionably, (a_F) and (a) have exactly the same semantics. They both encode the proposition *that Plato snores.* They both ascribe exactly the

same force to that proposition. Neither does anything besides this.

We may conclude that everything we just said about (a_F) applies exactly to (a). Let *t** be an arbitrary token of (a). *t** has for its semantic content a function that assigns the property of success to that same token exactly if a certain proposition is true. The proposition is: *that Plato snores.* *t** is, inter alia, a (phonetically distorted) token of "that Plato snores". A token of the latter denotes: *that Plato snores.*

*t** also comprises a (phonetically distorted) token of a force-operator. The force-operator assigns success to *t** exactly if *that Plato snores* is true. Thus, the semantic content of *t** is a function F that assigns the property of success to that token exactly if *that Plato snores* is true.

The semantic content of the sentence-*type* "Plato snores" is a function that assigns F to *t.*

Let us generalize these points.

Let *t* be *any* indicative sentence token. *t* will encode some proposition *p.* The semantic content of *t* is a function F that assigns success to *t* exactly if *p* is true.

Let *T* be the type of which *t* is a token. The semantic content of *T* is a function that assigns F to *t.*

The semantic content of an indicative sentence-type is a function that assigns such functions to tokens of that type.

The assertoric-force operator denotes a function that assigns the property of success to a token exactly if a certain proposition is true. A more exact definition of that operator must be given contextually. Given an expression token *t* that bears a proposition *p*, *t* comprises an assertoric force-operator exactly if some constituent of *t* denotes a function that assigns the property of success to *t* exactly if *p* is true.[291]

We will develop corresponding analyses for questions and imperatives.

This will involve defining the corresponding force-operators.

But before we do so, I would like to deal with a possible worry. It might seem that this account suffers from vicious circularity.

Let us consider a particular token of the expression type "Plato snores". Consider the one right below this line:

(*) "Plato snores".

That token is just an inscription – just a pattern of ink, not a pairing of ink with meaning.

Uncontroversially, there exists a function that assigns the property of success to that token just in case a certain proposition – *that Plato snores* -- is true. Obviously there is nothing wrong with that function; there is nothing circular about it. Let F* be that function.

To avoid vicious circularity, we must be careful how we state the theory. We don't want to say:

(A) The semantic content of an indicative sentence-token *t* is a function that assigns success to *t* if the proposition semantically encoded in *t* is true.

(A) is, or at least seems, circular. It defines the semantic content of *t* in terms of the proposition "semantically encoded in *t*". So it defines the semantic content of *t* in terms of the semantic content of *t.* Not acceptable.

What we must say is this:

(B) Part of what is encoded in an indicative sentence-token *t* is some proposition P. But that is not *all* that is thus encoded. *t* also comprises a

function that assigns success to that token just in case P is true.

There is no circularity in (B). *t* is just an arbitrary physical (or, at any rate, spatiotemporal) entity. For any proposition P, there is nothing wrong with saying that part of what *t* encodes is P. And for any function F, there is nothing wrong with saying that F assigns the property of success to *t.* That is not the kind of function we are used to considering; but that is obviously not important. It follows that there is nothing wrong with saying that F assigns the property of success to *t* exactly if P is true.

I should clarify one other point; this will head off another false accusation of circularity. Sometimes we say things like:

The occurrence of “can” can “canary” isn’t a word.

What we are saying is that a certain physical type is not appropriately paired with a meaning, and therefore doesn’t constitute a word (or an occurrence thereof). So, in that case, we are using the term “word” to denote a pairing of a physical object, or kind, with a meaning. Let us say that, in cases like this, we are using “word” in a “pregnant” manner.

We don’t always use “word” in that way. Sometimes we say things like:

The word “bank” has two different meanings.

Here we are saying that a certain physical type is paired with different meanings. So we are using the term “word” to denote a physical type, not a pairing of a physical type with a meaning. Let us say that, in such cases, we are using “word” in a “non-pregnant” manner.

Exactly similar remarks apply to most semantic terms: “sentence”,

“sentence-token”, “sentence-type”, “expression”.

In that analysis, I am using the term “sentence-token” in a non-pregnant manner. When I talk about a “token” of the type “Plato snores”, I am talking about a spatio-temporal object of some kind – an inscription, a noise, perhaps a mental image. I am not talking about a *pairing* of such an object with a meaning. If I were using the word “token” in a pregnant manner, then indeed my analysis would be circular. My analysis would then be:

(C) The content of an indicative sentence-token is a function that assigns success to that token under certain circumstances; and that sentence-token is a pairing of a physical object with a certain semantic content.

(C) defines “semantic content” in terms of “sentence-token”, and it then defines “sentence-token” in terms of “semantic content”. Obviously that is viciously circular. The reason it is circular is that “sentence-token” is being used in the pregnant manner.

In my analysis, I use it in the non-pregnant manner. So my analysis is not guilty of the circularity just described.

Questions and Imperatives

Questions and imperatives are to be analyzed in a similar fashion. Let us start with imperatives.

A token of an imperative denotes a function that assigns the property of success to that same token exactly if the addressee makes a certain proposition be true.

If I say: “march!”, my words are really elliptical for: “you: march!”. So if O is the addressee, then the proposition encoded in my utterance is: *that O*

marches. The “you” is semantically present, but phonetically unrealized.

To make things as clear as possible, I will always include the name of the person addressed. So I will not talk about utterances of “march!”, but about utterances of “Plato, march!”, “Socrates, march!”, and so on.

Suppose I am talking to Plato, and I say: “Plato, march!”. Let *t* be that token. Everyone who speaks English knows that, at the level of semantics, *t* is a success if it is obeyed and is a failure otherwise. Everyone who speaks English knows *which* proposition the addressee must make true if he is to obey that imperative. That proposition, of course, is: *that Plato marches.*

Again, we must distinguish semantic success and failure from other kinds. Many kinds of success can be achieved by uttering an unobeyed imperative: one can amuse, intimidate, manipulate, provoke, help, inform. And many kinds of failure can be incurred through issuing obeyed commands. But *semantically* an utterance of an imperative is a success exactly if it is obeyed. At some level, everybody who speaks a language knows that.

So, at the level of semantics, *t* is a success exactly if the addressee obeys it – exactly if, in response to my utterance, Plato makes true the proposition *that Plato marches.*

Thus, the semantic content of *t* assigns success to *t* exactly if, in response to *t*, the addressee makes true the proposition: *that Plato marches.*

So that semantic content is a function that assigns the property of success to that token exactly if, in response to that utterance, the addressee makes true that proposition.

One point should be clarified. Suppose I use an imperative to command somebody to do such and such. But suppose that person does such and such *not* because I commanded it, but for some totally unrelated reason. It seems to me that, in such a case, the imperative was a failure. This point is not central to my argument. But it seems true; and, I believe, it must be

accommodated. For an imperative to be a success, it is necessary that the addressee not only make true the proposition in question, but that he do so *in response* to the imperative.

Let us now state our analysis of imperatives in a more precise way. Unquestionably, an utterance of "Plato, march!" encodes the proposition: *that Plato marches.* This makes it reasonable to suppose that such an utterance contains a constituent that refers to that proposition. So, semantically if not phonetically, "Plato, march!" comprises an expression that refers to that proposition – presumably, "that Plato marches".

It is also clear that, in such an utterance, imperative force is being assigned to that expression. So "Plato, march!" is really a phonetically condensed or distorted form of:

(PM) "C(that Plato marches)",

where C is an operator that assigns imperative force to the expression falling in its scope. (The latter expression must denote a proposition.)

What exactly does C do? What is it for an expression to assign imperative force to an expression?

C is *not* merely a way of expressing the speaker's attitude towards a proposition. For the sentence

(**) "I want it to be the case that you, Plato, make true the proposition: *that Plato marches.*"

has a totally different semantics from (PM), even though it expresses specifically the propositional attitude just described.

For now familiar reasons, C does not replace one expression with another.

It is tempting to say:

An imperative force-operator replaces the proposition falling in its scope with some other proposition. The latter denotes the state of affairs that must obtain if the command is to be obeyed. So C is a function that takes (say) "that Plato marches" and replaces it with an expression denoting some second proposition. This second proposition is the one whose truth constitutes obedience of the command. This second proposition is: *that it is the case that Plato makes true the proposition* that Plato marches.

This proposal fails. Suppose that C is defined in the way just proposed. In that case, PM is synonymous with:

(**) "That Plato makes true the proposition *that Plato marches* (in response to my command)"

But (**) is not an imperative; it has no force, and is no synonym of PM.

Here is the right analysis of C. Let *t* be an arbitrary token of PM. *t* comprises a token of "that Plato marches". Of course, that expression denotes the proposition *that Plato marches.* C assigns the property of success to *t* exactly if, in response to *t*, the addressee makes true that proposition. Thus, the semantic content of *t* is a function F that assigns the property of success to *t* exactly if, in response to *t*, Plato makes true the proposition: *that Plato marches*

The semantic content of the *type* "C(that Plato marches)" is a function that assigns F to *t.*

PM is just a phonetically perspicuous version of "Plato, march". So

everything we said about the former applies to the latter. An utterance of "Plato, march" has for its semantic content a function that assigns success to that same utterance exactly if the proposition *that Plato marches* is made true by the addressee in response to that same utterance.

The semantic content of the *type* "Plato, march!" is a function that assigns functions, like the one just described, to each token of that type.

Let us generalize this analysis. Let *t* be any token of any imperative sentence-type. Part of what *t* semantically encodes is some proposition P. *t* semantically encodes, and thus refers to, a function that assigns success to *t* exactly if, in response to *t*, the addressee makes it be the case that P.

Questions

Questions are dealt with in basically the same way, but there are a few wrinkles.

First of all, there are two kinds of questions: those that are to be answered with a "yes" or a "no", and those that cannot be so answered. Let us refer to the latter as "ordinary" questions.

Yes-no questions always encode a complete proposition. A token of "does Plato snore?" encodes the proposition: *that Plato snores.*

Ordinary questions encode propositional functions: proposition-like entities that contain a free variable. "Where did Smith go?" encodes the propositional function: *That Smith went to__* There is a blank or free-variable. No complete proposition is encoded.

Let us start with yes-no questions. At the level of semantics, a question is a success only if it is answered correctly. If I ask a question, and receive no answer, then my query was a failure. This is a matter of semantics, of literal meaning; it is not a matter of pragmatics or anything else. Semantically,

questions are *supposed* to be answered. Semantically, a question is a success exactly if it is answered.

But a question is not just supposed to receive *an* answer: it is supposed to receive a *correct* one. Let Smith be somebody with a peculiar cognitive disorder. He knows that "does Plato snore?" is supposed to be answered; he also knows that an answer consists in an affirmation, or a denial, of the proposition *that Plato snores.* But Smith thinks that it is of no consequence whether the answer is correct or not; he thinks that the purpose of a question is simply to receive *an* answer, but not necessarily *the* answer.

It seems to me that Smith really doesn't understand the concept of a question. Smith really doesn't know quite how "does Plato snore?" is supposed to operate.

Further, this deficit on Smith's part is obviously semantic. It isn't that Smith has a perfect grasp of literal meaning, but falters when it comes to dealing with the nuances of implicature.

Suppose a student is irritatingly drumming his fingers on his desk, and you say, in a clearly annoyed way, "is it really necessary to do that ?" If the student is completely insensitive to pragmatic phenomena, but grasps literal meaning, he might say "technically, it isn't necessary; but I do feel like doing it, so I will carry on."

But if the student thinks that "does Plato snore?" is no more supposed to be given a correct answer than a false one, then we are no longer in the realm of implicature; we are dealing with a simple ignorance of semantics.

So a question q is a success only if it is given a correct answer. This is true both of yes-no and ordinary questions.

Of course, an unanswered question may be a great success in many respects, and an answered question may be a great failure in many respects. But *semantically* a question is a success only if it is answered.

Suppose I utter “does Plato snore?”. Let *t* be that token. Everybody who speaks English – who knows what that question means – knows that, in terms of literal meaning, that question is supposed to call forth a *true* a statement as to whether Plato snores. So every English-speaker knows that *t* is a success only if three conditions are satisfied. First, in response to *t*, the person addressed by that token correctly affirms or denies that a certain proposition is true. Second, that proposition must be either *that Plato snores* or *that Plato does not snore.* Third, the proposition affirmed must be true.

A couple of niceties must be dealt with. Suppose the addressee of *t* happens to produce a sentence that gives a correct answer to that question, but he does so for some reason having nothing to do with *t*; he does so by sheer coincidence. In that case, *t* is semantically a failure.

Suppose I say “is snow white?” to you. You say “snow is white”, but you are not saying this *in response* to my question; you are saying it for some totally unrelated reason – perhaps somebody previously agreed to pay you a million dollars if you said “snow is white” at that exact moment, regardless of what anyone else was saying. My question did not prompt you to say that. So my question was, at the level of semantics, a failure. It did receive an answer, but not for the right reason.

Questions are not only supposed to be answered: they are supposed to *prompt* people to answer them.

Here, then, is our analysis.

Let *t* be an arbitrary token of “does Plato snore?” At the level of semantics, *t* is a success exactly if, in response to *t*, the addressee correctly affirms or denies the proposition: *that Plato snores.* So the semantics of *t* assign the property of being a success to that token exactly if, in response to *t*, the addressee correctly affirms or denies that proposition.

Thus, the semantic content of *t* is a function F that assigns success to *t*

exactly if, in response to *t*, the addressee correctly affirms or denies the proposition: *that Plato snores*.

The semantic content of the expression-type "does Plato snore?" is a function that assigns F to *t.*

Let us generalize this point. Let *t* be any yes-no question-token you please. *t* semantically encodes some proposition P. The semantic content of *t* is a function F that assigns success to *t* exactly if, in response to *t,* the addressee correctly affirms or denies P. Let *T* be the corresponding type. The semantic content of *T* is a function that assigns F to *t.*

Ordinary Questions

Let us deal with ordinary questions. Let *t* be an arbitrary token of "where did Smith go?" *t* encodes not a proposition, but a propositional function, namely: *Smith went to place x.* Let P be that function. At the level of semantics, *t* is a success exactly if, in response to *t*, the addressee produces a sentence affirming a true proposition P* that is just like P, except that P* contains a place, where P contains a variable.

Let us say that P* is a "completion" of P.

So *t's* semantic content, its referent, is a function that assigns success to that same token exactly if, in response to *t*, the addressee affirms a true proposition of the kind just described.

Let us generalize this point. Let *t* be *any* token *any* ordinary question-type *Q*. *t* encodes some propositional function P. Semantically, *t* is a success exactly if, in response to *t*, the address affirms or denies a true proposition that is a completion of P. So *t*'s semantic content is a function F that assigns the property of success to *t* exactly if, in response to *t*, the addressee affirms or denies a true completion of P. The semantic content of Q is a function that

assigns F to *t.* It is obvious how to generalize this analysis to any ordinary question.

A Unified Semantics for Questions

The answer we've given to the question "what is the semantic content of a yes-no question?' isn't quite the same as the answer we've given to the question "what is the semantic content of an ordinary question?" But "question" is not a disjunctive category. So presumably, we should be able to give an answer that applies to both yes-no and ordinary questions.

This can be done. A proposition can be considered a limiting case of a propositional function: a propositional function with zero free-variables. This is, in fact, how most logicians think of propositions. So the proposition *that Plato snores* can be considered as a kind of propositional function and, by the same token, can be considered a completion *of* that same propositional function.

Let *t* be any token you please of any question-type you please. *t* semantically encodes some propositional function P. Semantically, *t* is a success exactly if, in response to *t* the addressee affirms a true completion of P. So the semantic content of *t*, its referent, is a function that assigns success to *t* exactly if, in response to *t*, the addressee affirms or denies a true completion of P.[292]

The General Definition of "Sentence"

The category of "sentence-token" is a unified one. Therefore, that category must be picked out by some one, non-disjunctive definition. What is that

definition? My answer is:

(TOK) *s* is a sentence-token exactly if, for some proposition P, *s*'s semantic content is a function F that assigns success to *s* exactly if P is true.

Let us now define "sentence-type". First of all, a sentence-type is a platonic entity. A platonic entity is something of which there are instances or tokens. Given this point, and given how we just defined "sentence-token", we must define "sentence-type" thus:

(TYP) *S* is a sentence-type exactly if, for any token *t* of S, there is some proposition P such that the semantic content of *S* is a function F that assigns a function f to *t* such that f assigns success to *t* exactly if P is true.

Obviously these definitions, especially (TYP), require elucidation.

Any sentence-token *s* – whether indicative, interrogative, or imperative – is such that, for some proposition P, s is a success exactly if P holds.

Consider a token of "Plato snores". Semantically, that token is a success exactly if: *that Plato snores* is true.

Consider a token of "Plato, snore!". Semantically, that token is a success exactly if Plato is the addressee and, in response to that token, he makes it be the case that: *that Plato snores* is true. So a token *t* of "Plato, snore!" is a success exactly if the following proposition is true: *Plato is the addressee of* t *and, in response to* t, *he makes it the case that:* that Plato snores *is true.*

Consider a token of "does Plato snore?". Semantically, that token is a

success exactly if, in response to that token, the addressee affirms a true proposition concerning whether *that Plato snores* is true. So a token *t* of "does Plato snore?" is a success exactly if the following proposition is true: *in response to* t, *the addressee affirms a true proposition concerning whether* that Plato snores *is true.*

It is obvious how to apply this analysis to an ordinary question, like "what did Plato eat?"

Given *any* sentence-token *s*, there is some proposition P such that *s* is a success exactly if P holds. So the semantic content of a sentence-token is a function that assigns it success exactly if some proposition holds. The semantic content of a sentence-type is a function that pairs its tokens off with functions like the one just described.

Of course, a thing does not *just* have semantic content; it has semantic content relative to the rules of this or that language. Thus, a thing *e* qualifies as a sentence-token only if, for some language L, the semantic rules of L assign a function F to e such that, for some proposition P, F assigns the property of success to *e* exactly if P is true. A thing *E* qualifies as a sentence-type only if it pairs off its tokens with functions like the one just described.

For reasons we have just seen, if a thing *e* is to qualify as a sentence-token, it is *necessary* that, as a matter of semantics, there be some proposition P such that *e*'s success turn on whether P is true. It seems to me that it is also sufficient. Let me say why.

How are sentences different from sub-sentential expressions? Wittgenstein and Dummett answer by saying: A sentence is a "complete move in the language game", whereas a sub-sentential expression is not. [293] There is probably truth in this, but it is much too vague. What is a "language-game"? What is a "move" in that game? Many think of a language as a set of functions: How is a set of functions a "game"?

A more perspicuous answer is not hard to produce. A sentence is something that, at the level of semantics, can be *successful*. A sub-sentential expression is something that, at the level of semantics, cannot be successful. If you say simply “Plato” or “because”, what you are saying cannot, at the level of semantics, be successful in any way. (Of course, I am considering tokens of “Plato” and “because” that are not elliptical for complete sentences. Obviously a token of “Plato” is successful if it is elliptical for a true sentence.) Given the semantics of English, there are no circumstances in which an isolated occurrence of “Plato” or “because” is a success. Semantically, any such occurrence is *always* a failure.[294]

But with *sentences* things are very different. Let *S* be any sentence-token of English. There is some circumstance C such that, given the semantic rules of English, S is a success exactly if C obtains.

As a matter of semantics, there is some circumstance C such that “Plato snored” is a success exactly if C obtained. Here C is the circumstance that Plato snored.

Of course, logically speaking, there is *no* circumstance C such that “triangles have four sides” is a success if C obtains. But *semantically* speaking, there is such a circumstance. The *semantics* of English leave it open whether “triangles have four sides” can be a success or not: what prohibits success is a fact of mathematics, not of semantics. The rules of English semantics *do* assign a success-condition to that sentence: so, from the viewpoint of semantics, *there is* some circumstance C such that “triangles have four sides” is a success exactly if C obtains. That is why we can say that any such token is necessarily false: we know of the necessary non-existence of the condition under which such a token is assigned success by the semantic rules of English. So from the standpoint of English semantics, *there is* a condition under which tokens of “triangles have four sides”.

The same is true of any question or imperative. A token *t* of "shut the door", addressed to O, is a success exactly if, in response to that token, O shuts the door. So *t* is a success exactly if the following proposition is true: *that, in response to* t, *O shuts the door.*

Of course, some commands cannot be obeyed. But, from the viewpoint of English semantic rules, given any imperative *t*, there is some condition C such that *t* is a success exactly if *t* is obeyed. Any token *t* of "destroy an indestructible object" cannot be obeyed. But, from the viewpoint of English semantics (though not of logic), there is some circumstance C such that *t* is a success exactly if C obtains. That circumstance is that the addressee, in response to *t*, destroy some object that cannot be destroyed. Of course, that cannot be done. But that impossibility is not *semantic* in nature. From the viewpoint of English semantics, *there is* a circumstance in which that command is obeyed and in which, thus, it is a success.

By contrast, given any *sub*-sentential expression *s*, there is *no* circumstance under which, from the viewpoint of English semantic rules, *s* is a success. (Again, I am obviously assuming that *s* is not elliptical for a complete sentence.) Of course, saying "is" in isolation might bring one great success. (Maybe someone you are wooing likes the way it sounds.) But that has nothing to do with semantics.

So it is not so wrong to define "sentence-token" thus: *S* is a sentence-token of a language L exactly if there is some circumstance C such that a semantic rule of L assigns the property of success to *S* exactly if C holds. For C to hold is for some proposition to be true. So *S* is a sentence-token exactly if, as a matter of semantics, its semantics turns on the truth of some proposition – exactly if its semantic content is a function that assigns it success just in case some proposition holds. *S** is a sentence-type exactly if it assigns functions like the one just described to its tokens.

Let us sum up. *S* is a sentence-token if, semantically, it has success-conditions. *S** is a sentence-type if, at the level of semantics, it assigns success-conditions to its tokens.

A Word of Caution

Our analyses of "sentence-type" and "sentence-token" are rich in quantifiers. It is very important to put the quantifiers in the right order. Consider:

(TYP) *S* is a sentence-type exactly if, for any token *t* of S, there is some proposition P such that the semantic content of *S* is a function F that assigns a function f to *t* such that f assigns success to *t* exactly if P is true.

(TYP) must be distinguished from the falsity:

(TYP_F) *S* is a sentence-type exactly if, for some proposition P, for any token *t* of S, the semantic content of *S* is a function F that assigns a function f to *t* such that f assigns success to *t* exactly if P is true.

Consider the sentence-type "I am tired". Thus there is no proposition P such that, for any token *t* of "I am tired", *t*'s success turns on the truth of P. When you token "I am tired" the success of that token turns on one proposition; when I token it, the success of *that* token turns on a different proposition. So there is no proposition P such that the type "I am tired" assigns a function F such that, for any token *t* of "I am tired", F assigns a function f to *t* exactly if P is true.

The quantifier “for some proposition P” must not come before the quantifier “for any token *t*”.

If we put “for some proposition P” *after* “for any token *t*”, then our definition is correct. Given any token *t* of “I am tired”, there *is* some proposition P such that *t* is a success exactly if P is true. The semantic content of the *type* “I am tired” is a function F that pairs off its tokens with functions like the one just described. Thus the semantic content of “I am tired” is a function F such that, for any token *t* of that type, there is some proposition P such that F assigns a function f to *t* exactly if P is true.

A Desideratum

Many think it absurd to suppose that sentence-tokens refer to anything. One reason for this, I feel, is that (indicative) sentence-tokens are true or false, while paradigmatic referring terms are not. “Plato” is not true or false. Expressions referring to propositions – for example, “that Plato snores” – are not true or false, even though the things they denote are true or false.

Sentence-tokens are true or false; paradigm cases of referring terms are neither. It follows, we are told, that sentence-tokens don’t refer.

This argument is confused; it involves a wrong conception of reference. It involves the erroneous idea that reference is severable from semantic contribution. But that argument points to a desideratum that any correct semantics for sentences must satisfy.

We observed that a *certain* circularity characterizes my analysis of sentence-tokens and sentence–types. Any analysis that satisfies the desideratum just mentioned will be characterized by a similar circularity.

Some say that sentences refer to truth-values. But it is very easy to produce an expression that uncontroversially refers to a truth-value but lacks

force, and therefore isn't a sentence at all:

(t) "the truth-value *true*".

If you agree with Russell's theory of descriptions, then consider "ze truth-value *true*" or "dthat [the truth-value true]".

Let

(s) "snow is white"

be any token of "snow is white".

(t) doesn't have force; it isn't true or false, even though what it *refers* to is true or false.

Of course, those who say that (s) refers to a truth-value will say:

What is true or false is the *sense* of (s), not its referent.

This move is a failure. Types are what have sense, not tokens; tokens are what refer, not types. Tokens refer to the things they are assigned by the corresponding types; and tokens have no semantic content *other* than what they are thus assigned. So if we say that (s) refers to the truth-value *true*, then we are saying that it has no *other* semantic content. Thus (s) has exactly the same semantic content as a token of "the truth-value *true*". The latter is neither true nor false. So it must have a different semantic content from (s). So the idea that sentence-tokens refer to truth-values doesn't do justice to the fact that sentence-tokens are true or false.

This problem applies, though less acutely, to the theory that sentence-

tokens denote propositions. Given any proposition, we can produce an expression that is neither true nor false, and that has that proposition for its sole semantic content. Let

(s*) “that snow is white”

be a particular token of “that snow is white”.

Obviously (s*) refers to proposition *that snow is white.* But (s*) isn’t true or false. So it doesn’t have the same semantics as (s).

Granted, (s*) and (s) *almost* have the same semantics. They both semantically encode *that snow is white*; and neither semantically encodes any other proposition. So it is *largely* right to say that (s) and (s*) refer to the same thing. But it isn’t quite right.

Any theory that purports to give the semantics of indicative sentence-tokens must explain why such tokens are true or false. If we say that sentence-tokens have such and such semantic content, it is absolutely crucial that no expression having such and such content can *lack* assertoric force.

The existing theories – sentence-tokens refer to truth-values, sentence-tokens refer to propositions -- miserably fail this test.

Barwise and Perry[295] say that the semantic content of a sentence-token is a “situation” or a “class of situations.” They identify propositions with classes of situations. I think there is profound truth in their position. For reasons we considered, I think that sentence-tokens *do* refer to propositions. But that is not *all* they refer to. This means that the Barwise-Perry position must be extended if it is to be completely accurate. Perhaps this will explain my meaning.

Let C be any situation, or class of situations, you please. We can produce an expression that is *not* a sentence-token, and whose referent, and whose

sole semantic content, is C. Let “SC” be a name for C. “SC” isn’t any more true or false than “Caesar”. Of course, sentence-tokens typically are true or false. Therefore the semantic content of a sentence-token is not *just* a proposition (though that is part of what it is); therefore (supposing that Barwise and Perry are right to identify propositions with classes of situation), the content of a sentence-token is not *just* a class of situations.

Most expressions referring to functions are not true or false. This fact is especially important. For it is tempting to analyze sentence-tokens as follows. The semantic content of a sentence-token is a function that assigns *truth* to that sentence-token exactly if a certain proposition is true. So the semantic content of (s) is some function F that assigns the property of *truth* to (s) exactly if *that snow is white* is true.

The problem is: we can easily produce an expression whose sole semantic content is F, and that isn’t true or false. Here is an example. Let “F” be an expression whose sole semantic content is F. “F” is not true or false; it is no more a sentence than “Caesar”. So the semantic content of “snow is white” cannot be F, at least not F alone.

Nothing that lacks assertoric force can have the same semantic content as (s), or any other sentence-token. There may be overlap, but not identity, of semantic content. Hence, the following is a desideratum that any semantic theory must satisfy:

(D) The semantic content of an indicative sentence-token must be something that an expression could not have *without* being true or false.

Any function assigning truth or falsity to (s) can be the sole semantic content of an expression that has no assertoric force. Any proposition, any

fact, any truth-value, any mental state, any physical object: these can be the sole semantic content of an expression that has no assertoric force.

This is what motivates our theory. At the level of semantics, of literal meaning, (s) is a success exactly if a certain proposition is true. So (s)'s semantic content, its literal meaning, must at least include a function that assigns the property of success to (s) exactly if that proposition is true.

The semantic content of an indicative sentence-token must be a function that assigns that very token the property of success exactly if a certain proposition is true.

This theory, and no other, satisfies (D).

Let *t* be *any* expression-token you please. Suppose that *t*'s semantic content is a function phi that assigns success to *t* exactly if some proposition P is true. In that case, there is no way that *t* could fail to be an indicative sentence.

By assigning phi to *t*, you've guaranteed that, at the level of literal meaning, *t* is a success if P is true. You've guaranteed that *t* is failure if P is false. You've made it a matter of *t*'s semantics that *t* is a success just in case P is true. Thus, *t*'s semantic success turns on the truth of P. *t* asserts that P.

If you don't believe this, do an experiment. Consider the proposition *Kennedy's favorite author was Tolstoy.* (I chose this proposition because I do not know whether it is true or false.) Let "K" be some particular expression-token. So "K" is some particular bit of sound or ink or mental imagery.

Let us stipulate that "K" has a certain F* function for its semantics. F* is defined thus: F* assigns the property of success to "K" exactly if the proposition *that Kennedy's favorite author was Tolstoy* is true.

So *semantically* one thing is necessary and sufficient for "K" to be a success: the truth of the proposition *that Kennedy's favorite author was Tolstoy.*

How do we go about evaluating whether, at the level of semantics, “K” is a success? We find out if Kennedy’s favorite author was Tolstoy. That is the only thing we do.

So “K” doesn’t just encode the proposition *that Kennedy’s favorite author* was Tolstoy. It *does* do that. But it does more. Semantically, “K” is a success exactly if that proposition is *true.*

Anyone who speaks English, and knows the semantics of “K”, knows that, in terms of literal meaning, “K” and a token of

(k) “Kennedy’s favorite author was Tolstoy”

are successful exactly if *that Kennedy’s favorite author was Tolstoy* is true.

So our analysis correctly predicts that any token of (i) will be synonymous with “K”. Our analysis makes it *impossible* for a sentence-token to have for its semantic content anything that could be the semantic content of an expression that was not true or false.

Every other analysis fails to account for this, even the theories – *especially* the theories – that identify the semantic content of sentence-tokens with the properties of being true or false. For plainly an expression can have the property of being true for its sole semantic content and not be true (or false). Consider, for example: “the property of being true”.

Chapter 13 Our Semantics

Given what we’ve said in this work, certain reasonable-seeming semantic views must be adjusted.

In order to give a unified view of our semantic results, we must first deal with one problem. Part of what we are trying to do is to give an analysis of noun-phrases. Among such phrases are different kinds of existential quantifiers: "something", "some man", "three birds". According to Russell, definite descriptions are existential quantifiers.

Frege has given us a very plausible analysis of such expressions. But a commonly accepted analysis of them actually makes Frege's analysis be wrong; more exactly, it makes Frege's analysis be only *partially* correct. I believe that Frege's analysis is unqualifiedly correct. So I must show why this other analysis, due to Quine[296], is wrong.

Quine on "some"

According to Frege, "some" and "for some x", and the like, are to be defined contextually:

(F) "…some phi…" means: *there is some x such that x has phi and…x…*

In some cases, the expression "a phi" is to be analyzed in the same way. Sometimes "a" means *any*: "a man is only as good as his word". But sometimes it means "some": "a man stole my wallet". We will focus on the second meaning.

As Quine[297] pointed out, some sentences of the form "…a phi…" (or "…some phi…") are ambiguous between wide-scope and narrow-scope readings. On the one hand,

(i) "John wants a sloop"

can be taken to mean that there is some *particular* sloop that John wants. On the other hand, (i) can be taken to mean that John wants mere "relief from slooplesness", as Quine put it.[298]

An exactly similar ambiguity is generated whenever an occurrence of "some" falls within the scope of an epistemic operator like "believes" or "is looking for". Let us initially discuss "wants"; we will then generalize our analysis to these other operators.

According to Quine, the first reading of (i) is this.

(i_W) There is some x such that x is a sloop and John wants x.

The second reading, says Quine, is this:

(i_N) John wants: that there is some sloop x such that John has x.

I have only one disagreement with Quine. Notice that in (i_W), the complement of "wants" is an *object*, whereas in (i_N) the complement of "wants" is a *proposition.* If that is right, then in some cases what is wanted is a *thing* – a vase, a sloop, a car -- whereas in other cases what is wanted is that a certain *proposition* be true: one involving a vase, a sloop, a car. So on Quine's analysis, "wants" is ambiguous between "wants$_T$" and "wants$_P$". The former takes objects for its complements; the latter takes propositions. The latter denotes a propositional attitude; the former does not – it denotes an "objectual attitude".

I doubt that "wants" is ambiguous, semantically or even just

pragmatically, between "wants$_T$" and "wants$_P$". I am pretty sure that the kind of thing "wants" denotes in (i_W) is identical with the thing it denotes in (i_N). I am pretty sure that "wants" *always* denotes a propositional attitude.

Also, even if we leave aside matters of semantics, I really doubt that the very concept of desire is thus disjunctive: I doubt that there is one kind of wanting that concerns states of affairs and another that concerns objects. I may "want" a particular sloop; I may "want" *some* sloop or other, not this or that specific sloop; but what varies is not the nature of wanting, but the identity of the thing I want.

Here is what I would propose. In *every* case, the object of "wants" is a proposition. When we say that so and so and "wants" x, we are really saying that so and so wants some proposition to be true concerning so and so's relation to x. If I say: "Smith wants a car", I am saying: Smith wants it to be the case that: Smith has a car.

The story is not relevantly different if there is some *specific* car that Smith wants. Suppose we say: " there is some *particular* car x such that Smith wants x." What we are saying is: " there is some *particular* car x such that Smith wants that: Smith has x."[299]

So Quine's *narrow*-scope reading of (i) is quite correct. It is his *wide*-scope reading that is wrong. The right wide-scope reading of (i) is not (i_W). It is:

(ii) There is some particular sloop x such that John wants: that John has x.

I would like to briefly discuss the difference between *semantic* and *syntactic* ambiguity. Any adequate semantics for intensional verbs *requires*

them to *be syntactically* [300] ambiguous. At the same time any such semantics must not make them *semantically* ambiguous. Quine's analysis makes them semantically, and not just syntactically, ambiguous.

What do these terms mean?

Every student of logic is familiar with the sentence:

(a) Every man loves a woman.

There are two readings:

(b) There is some woman x such that, for any man y, y loves x.

(c) Given any man x, there is some woman y, such that x loves y.

(b) says that some particular woman is loved by all men. (c) does not say this.

So (a) is ambiguous. But it is not *semantically* ambiguous. No constituent has two readings. Each constituent has one reading, but these constituents can parse out in different ways.

Another famous example of syntactic ambiguity is:

(d) "it is not likely that any given person will win the lottery".

This is ambiguous between

(e) For any x, it not likely that x will win the lottery

and

(f) It is not likely that, for any x, x will win the lottery.

(e) is true; (f) is false. (d) is ambiguous. None of its constituents has more than one reading. But those constituents can parse out differently; so (d) is syntactically ambiguous.[301]

In light of this, consider the sentence-token:

(i) “John wants a sloop”.

Quine says that this is ambiguous between:

(i_W) There is some x such that x is a sloop and John wants: x.

and

(i_N) John wants: that there is some sloop x such that John has x.

(i_N) attributes a *propositional attitude* to John. (i_W) does *not* attribute any propositional attitude to John; it attributes an "objectual" attitude to John - an attitude towards an object.

There are two reasons this is not acceptable. First, if you say that John "wants" anything, you are *always* attributing a propositional attitude to him, even if we give a de re reading to the statement in question.

Suppose John wants some *specific* sloop – say, *The Winds of War*. What John wants is that a certain *proposition* be true, namely: the proposition *that John owns The Winds of War*.

Even if you want some *specific* woman – Sally, say -- to be your bride, you still have an attitude towards a proposition; you want the proposition *that Sally is my bride* to be true. You don't just want: *Sally*.

In general, even when it is a *specific* thing that one wants, one has a propositional, not an objectual, attitude.

That by itself shows that Quine's analysis is wrong. But even if we waive the last point, Quine's analysis still fails. If "wants" sometimes denoted a propositional attitude, and sometimes did not, then "wants" would surely be *semantically* ambiguous: how could an expression that denoted two entirely different things fail to be? But "wants" is not semantically ambiguous, at least not in the way demanded by Quine's theory.

For reasons exactly similar to those just stated, Quine's analysis makes *any* intensional verb be semantically, and not just syntactically, ambiguous. It is pretty clear that his de dicto reading is right. That reading does justice to the fact that, when John wants something, he has an attitude towards a proposition. It is Quine's de re reading that is wrong.

We must replace Quine's de re reading with the one we have proposed:

(ii) There is some particular sloop x such that John wants: that John has x.

This reading does justice to the fact that, whenever John wants something, he has a propositional attitude. Also, unlike Quine's reading, our reading does *not* render "wants" semantically ambiguous. Our "wants" always denotes the same propositional attitude.

Of course, the specific proposition that John wants to be true will vary. Sometimes he will want the truth of: *John owns some boat or other*. Sometimes John will want the truth of: *John owns the Queen Mary*. But "wants" always denotes a propositional attitude; it doesn't shift between objectual and propositional meanings.

Analogues of Quine's analysis render *all* intensional verbs semantically ambiguous. Analogues of the analysis we just gave appropriately undo that ambiguation.

the relevance of this to our project

According to Frege, (F) is the semantic rule for "some phi". But it *seems* that when we apply (F) to (i), what we produce is never (i_N); it *seems* that, to generate (i_N), we cannot apply (F) in a purely mechanical way, but must make *ad hoc* adjustments. This would suggest that (F) is not quite accurate.

At the same time, (i_N) is, I have no doubt, a legitimate reading of (i). Obviously not *all* legitimate readings of (i) are de re readings. Obviously intensional verbs are often to be given de dicto readings.

My point is that none of this is any threat to the correctness of (F). (F) is correct. Applied in a purely mechanical way, without any *ac hoc* adjustments being made, (F) always yields the right result when applied to (i), or to other sentences containing existential quantifiers falling in the scope of intensional verbs. To see this, we must realize that (i) is really a compressed form of:

(i_C) "John wants: that John has a sloop.

(i_C) is indeed ambiguous. But (F) generates exactly those two dismambiguations and no others. When we apply F, what results is either:

(ii_{CN}) John wants: that there is some x such that x is a sloop and John has x

or

(ii_{CW}) There is some sloop x such that John wants: that John has x.

Both readings are generated by (F). The way that Quine reads (i) makes (F) turn out false, at least on one legitimate reading. Quine's reading has the consequence that (F) is right for the wide-scope/de re reading, but wrong for the narrow-scope/de dicto reading. But (F) is right for both readings. We must simply remember that "wants", and other intensional verbs, unfailingly have *propositional* complements; and we must remember that phonetics sometimes obscures this fact.

So before we apply (F) to (i), we must undo the distortions of phonetics. We must rewrite (i) thus:

(i_P) “John wants: that John has a sloop”.

When we apply (F) to (i_P), we generate two readings: and *both* are correct:

(i_{PN}) “John wants: that there is a sloop x such that John has x.”
(i_{PW}) “There is a sloop x such that John wants: that John has x.

Thus (F) gives the right reading across the board.

As we saw earlier, if (F) is right, then “some” (and “a”) can be seen as denoting a function that assigns truth to a class exactly if that class is instantiated. We can thus accept that definition, and know that no modification of it is demanded by contexts like (i).

We saw how Quine’s analysis partially falsified Frege’s treatment of “some phi”; it makes Frege’s analysis produce the right reading in only *some* cases. If we extend Quine’s analysis to other quantifiers, it yields exactly parallel falsifications of the time-tested analyses of them as well. Consider the sentence

(b) “John wants every boat”.

If we apply Quine’s analysis to (b), we end up with:

(b_N) John wants: that for any boat x, John has x.

and

(b_W) For any boat x, John wants: x.

If this natural extension of Quine's analysis is correct, then "wants" denotes a propositional attitude in the de dicto reading of (b); and in the de re reading, it denotes something that is *not* a propositional attitude – it denotes an "objectual" attitude. But this is absurd. Obviously "wants" always denotes a propositional attitude. So Quine's analysis fails.

More importantly, that analysis falsifies the standard analysis of "every phi". The standard analysis is:

(e) "…every phi…" means that: for any x such that x has phi…x….

(e) is an algorithm: a mechanical way of assigning meaning to sentences of the form "…every phi…" If Quine's analysis of (b) is correct, then only *one* reading of (b) is generated by (e). That reading is (b_W). The problem is that (b_W) is not even *a* legitimate reading of (b); for in (b_W) the verb "wants" doesn't denote a propositional attitude. So Quine's analysis, extended to "every", makes (e) be quite false.

But (e) is a time-tested and fruitful analysis. If Quine's analysis were right, we would have to make drastic changes to our current semantic theories.

Fortunately, we don't have to do this. What is false is not (e), but Quine's analysis. The verb "wants" always has a propositional complement. The same is true of any intensional verb. Before we apply (e) to a sentence containing such a verb, we must take care to make its semantics coincide with its phonetics:

(b) John wants: that John has every boat.

If we apply (e) in a completely mechanical way to (b), we generate two readings:

(b_{WK}) For any boat x, John wants: that John has x.
(b_{NK}) John wants: that for any boat x, John has x.

On this analysis, "wants" denotes a propositional attitude on the de re reading: this is as it should be, since "wants" *always* denotes such an attitude. Further, "wants" denotes the very same propositional attitude on the de dicto reading. Again, this is as it should be: for "wants" is not semantically ambiguous, at least not in (b).

Of course, (b_{WK}) and (b_{NK}) ascribe different propositional attitudes to John. But this is not because "wants" has different meanings in them. It is because different propositional functions are the verbal objects of "wants". Quine's analysis wrongly has "wants" denoting altogether different kinds of thing.

Most important of all, on our analysis, (e) is unqualifiedly right: it *always* gives the right meaning.

If (e) is right, then "every phi" denotes a function that assigns truth to a class C exactly if C is universally instantiated.

Since (e) is right, we can accept that analysis.

What we've said about (e) and (F) is true of many other quantifiers: "three men", "two girls", "most rabbits". There are time-tested and fruitful analyses of these quantifiers. When we extend Quine's analysis to occurrences of them that fall within the scope of intensional verbs, these analyses turn out to be falsified, for reasons exactly parallel to those given above. We undo this damage by rejecting these extensions of Quine's analysis, and thus by

rejecting Quine's original analysis; this involves realizing that intensional verbs always take propositional complements, even when given de re readings.

our semantics

Now we can give a unified semantics for noun-phrases.

We say that expression-types don't refer to anything. So the type "for some x" doesn't refer to anything, not even a function. But tokens of it do.

So if *t* is a token of "for some x", then the semantic content of *t* is a function F that assigns truth to a class C exactly if C is instantiated.

Let *T* be the type corresponding to *t.* The semantic content of *t* is a function that assigns F to *t.*

Exactly similar remarks apply to "some" and (mutatis mutandis) to "some man", "something", and so on.

It is pretty clear how this analysis generalizes to other quantifiers. If *t* is a token of "nothing" ("everything", "most things", "many things"), then *t* denotes a function F that assigns truth to a class C exactly if C is uninstantiated (universally instantiated, usually instantiated, instantiated to a non-negligible degree). The semantic content of *T*, the corresponding type, is a function that assigns F to *t.*

Obviously "some man" is in the same *grammatical* category as "Plato" and "that man", "the author of Waverly", and "ze author of Waverly".

Intuitively, we want to say that there is some significant and non-disjunctive property had in common by all and only those things we call "noun-phrases". Semantics has to respect grammar, at least up to a point. Grammatical categories may be useless when it comes to logic or metaphysics. But they cannot be useless when it comes to semantics. How

could they be? When you say that x, y, and z are “adjectives” or “determiners” or “nouns”, you are saying something specifically about the kinds of contributions that x, y, and z can make to the meanings of sentences in which they occur; you are saying something specifically about the *semantics* of such expressions. So it is a truism that semantics, at least the semantics of natural language, must respect grammar.

We don’t want to say that (tokens of) *some* noun-phrases denote second-level functions, while others denote individuals. That, it seems to me, would make “Plato” fall into a very different grammatical category from “some man” and “that man”.

This last point, I grant, is not exactly air-tight. But what is clear is that *if* “Plato”, “some man”, “all birds” all denote the same kind of thing – if they all denote individuals or they *all* denote functions – then it will be a lot easier to do justice to the fact that they are all in the same grammatical category than it would be if we said that some of them denote an individual, while others denote, not just functions, but second-order functions (functions that take functions as arguments).

The problem is that “some man” does not denote *any* individual. No quantifier does. There is no doubt about this. Inevitably, some say that “some man” denotes an arbitrary individual[302] or an individual in somebody’s “belief-world”.[303] But Russell’s (1920) arguments against this view seem decisive.

At the same time, we know from Kripke (1972) that “Plato” is not the kind of quantifier that Russell thought. It is hard to believe that “Plato” is anything other than a term that contributes Plato. Thus, “Plato” denotes Plato, and not a function: and once again we have the problem of giving a unified semantics for “Plato”, “some man”, and so on.

I think there may be a solution.

A token of “Plato snores” is true exactly if the class of snorers comprises Plato. So “Plato” can be seen as denoting a function that assigns truth to a class exactly if Plato is a member of that class.

Thus a token of “Plato” denotes a function that assigns truth to a class C exactly if Plato is a member of C.

Tokens of “Plato” are thus in the same basic category as tokens of “some man”, “no man”, and so on. They all denote functions that assign truth-values to classes.

Tokens of “Plato” are not exactly *quantifiers.* A quantifier is an expression that denotes a function assigning a truth-value to a class on the basis of *how many* things of a certain type are in that class. That is not what tokens of “Plato” do.

But that doesn’t threaten the semantic unity we have established. All of “Plato”, “some man”, and so on, denote the same basic kind of function.

Our analysis is consistent with Kripke’s insights. On our analysis, “Plato snores” is true exactly if *Plato* snores – it isn’t necessary that a teacher of Socrates or an author of the *Republic* be a snorer. So tokens of “Plato” are directly referential. But they are directly referential by way of encoding functions of the kind just described.

The semantic content of the expression-type “Plato” is a function that assigns the functions just described to tokens of “Plato”.

We know from Kaplan (1989) that tokens of indexicals are directly referential. For example, a token of “you”, addressed to so and so, semantically contributes so and so himself, as opposed to a concept or description that applies to so and so. If I address Smith, and I say “you are looking well today”, the proposition encoded in my words is: *Smith is looking well today.* The token of “you” is *directly referential*: it contributes Smith,

not a description that applies to Smith. In general, tokens of “you” are directly referential.

In my view, this is correct. But it must be adjusted. We want such tokens to be in the same general category as “Plato” and “some man”. So the right rule is this. A token of *t* “you”, addressed to x, denotes a function F that assigns truth to a class exactly if x is a member of that class. If there is no addressee, the utterance is abortive. The semantic content of *T*, the corresponding type, is a function that assigns F to *t.*

If Russell is right about definite descriptions, then for any token *t* of “the phi”, *t* denotes a function F such that, for any class C, F assigns truth to C exactly if for some x, x is a phi that falls in C, and no phi falls outside of C. The type *T* is the function that assigns F to *t.* (I am ignoring niceties relating to contextual salience.)

If the referentialist is right, then “the phi” is defined thus. If it is the case that, for some x, x is a unique phi, then a token *t* of “the phi” denotes a function F that assigns truth to a class exactly if C comprises x. If there is no such x, then *t* denotes nothing – “the phi” is abortive in that case. The semantic content of *T*, the corresponding type, is a function that assigns F to *t.*

The last two paragraphs are *virtually* indistinguishable. The only difference lies in the order in which the quantifiers are presented. The referentialist position gives wide-scope to the “for some x”: it is the first quantifier in the sentence that gives his analysis. The Russellian position gives that quantifier much narrower-scope; it occurs within the scope of several other quantifiers. Because of this difference, the referentialist has “…the inventor of bifocals…” be *de re* about Franklin, whereas the Russellian has such sentences *not* be de re about him. But notice how slight the difference is between these two analyses: this corresponds to the fact that, contrary to what

is commonly thought, the referentialist position *is* consistent with the cognitive significances of sentences containing definite descriptions.

The character of a sentence-type containing an indexical is not a function from contexts to propositions. The semantic content of "I am tired" is a function that assigns meaning to its tokens. If x uniquely utters a token *t* of "I am tired", the meaning of *t* is a function F that assigns success to *t* exactly if x is tired. The semantic content of the type "I am tired" is a function that assigns F to *t.*

The semantics just outlined is obviously Montague-inspired. But it makes some changes to Montague's semantics, corresponding to the developments in semantics that have occurred since the time of his writing. Our analysis accommodates the fact that tokens of proper names and indexicals are directly referential. It also accommodates the fact that expression-types have a very different semantic content from expression-tokens, even where proper names, and other apparently one-dimensional expressions, are concerned.

Bibliography

Austin, J.L. 1962. *How to Do Things with Words*. Cambridge: Harvard University Press.

Bach, Kent. 1984. *Thought and Reference.* Oxford: Clarendon.

Barwise, Jon and Perry, John.1990/1983. "Semantic Innocence and Uncompromising Situations". In *The Philosophy of Language.* (1990.)

Second Edition A.P. Martinich (ed.).
Oxford: Oxford University Press.
-- *Situations and Attitudes.* 1999. Palo Alto: CSLI Publications.

Blackburn, Simon. 1984. *Spreading the Word.* Oxford: Clarendon.

Brandom, Robert. 1998. *Making it Explicit.* Cambridge: Harvard University Press.

Carnap, Rudolph. 1934. *The Logical Syntax of Language.*
--1947. *Meaning and Necessity.* Chicago: University of Chicago Press.
-- 1958. *Introduction to Symbolic Logic and its Applications.* New York: Dover Publications.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge: M.I.T. Press.
--1975. *Reflections on Language.* Reprinted in *On Language* (1998). New York: St. Martin's Press.

Church, Alonzo. "Carnap's Introduction to Semantics". *The Philosophical Review.* Vol. LII. No. 3. 1943. 298-304.
--1956. *Introduction to Mathematical Logic.* Princeton: Princeton University Press.

Davidson, Donald. 1984. "True to the Facts". In *Essays on Truth and Interpretation.* Oxford: Oxford University Press.

--1965. “Theories of Meaning and Learnable Languages”. In *Essays on Truth and Interpretation.* Oxford: Oxford University Press.
--1967. “Truth and Meaning”. In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.
--1979. “Moods and Performances”. In *Essays on Truth and Interpretation.* Oxford: Oxford University Press.
Donnellan, Keith. 1966. “Reference and Definite Descriptions.” In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.

Dummett, Michael 1981. *Frege: Philosophy of Language.* Second Edition Cambridge: Harvard University Press.

Evans, Gareth. 1990/1974. “The Causal Theory of Names”. In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.
-- 1980. “Pronouns”. *Linguistic Inquiry* 11: 337-362.
-- 1982. *The Varieties of Reference.* Edited by John MacDowell. Clarendon: Oxford University Press.

Fodor, Jerry. 1990a “Fodor’s Guide to Mental Representation”. In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
-- 1990b. “A Theory of Content I”. In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
--1990c. “A Theory of Content II”. In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.
--1990d. “Substitution Arguments and the Individuation of Beliefs”. In *A Theory of Content and Other Essays.* Cambridge. The M.I.T. Press.

--1998. *Concepts.* Clarendon Press: Oxford.

Frege, Gottlob. 1879. “Begriffsschrift”. In *The Frege Reader.* Edited by Michael Beaney. Oxford: Blackwell.
1997/1891. “On Function and Concept.” In *The Frege Reader.* Edited by Michael Beaney. Oxford: Blackwell.
--1892a. “On Sense and Nominatum.” In A.P. Martinich (ed) *The Philosophy of Language* Oxford: Oxford University Press.
-- 1892b. “On Concept and Object.” In *The Frege Reader.* Edited by Michael Beaney. Oxford: Blackwell.
-- 1884. *The Foundations of Arithmetic.* Translated by J.L. Austin. Oxford: Basil Blackwell. Reprinted in 1980 by Northwestern University Press.

Geach, Peter. 1962. *Reference and Generality.* Ithica: Cornell University Press.
-- 1965. “Assertion”. *Philosophical Review*, 74 (1965), 449-465.

Gödel, Kurt. 1944. “Russell’s Mathematical Logic.” In *The Philosophy of Bertrand Russell.* Edited by Paul Arthur Schilpp. Evanston, Illinois: Northwestern University Press.

Kaplan, David 1964. *Foundations of Intensional Logic.* PhD dissertation. University of California, Los Angeles.
--1990/1970. “Dthat”. In A.P. Martinich (ed). *The Philosophy of Language.* Oxford: Oxford University Press.
Kaplan, David. 1978. “On the Logic of Demonstratives”. *Journal of Philosophical Logic* 8 (1978): 81-98.

-- 1989. ‘Demonstratives’. In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.
-- 1989b. “Afterthoughts”. In *Themes from Kaplan.* Edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.

Kripke, Saul. 1971. “Identity and Necessity”. In *Metaphysics: An Anthology.* Edited by Jaegwon Kim and Ernest Sosa. Oxford: Blackwell.
--1972. *Naming and Necessity.* Cambridge: Harvard University Press.
-- 1977 ‘Speakers Reference and Semantic Reference’. In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.

Lewis, Clarence. 1946. *An Analysis of Knowledge and Valuation.* Cambridge: Harvard University Press.

Lewis, David. 1972. “General Semantics”. In *Semantics for Natural language.* Edited by Donald Davidson and Gilbert Harman. Dordrecht-Holland: Reidel.
--1984. *On the Plurality of Worlds.* Oxford: Blackwell.

McCawley, James D. 1981. *Everything that Linguists Have Always Wanted to Know about Logic*. Chicago: University of Chicago Press.
--1998. *The Syntactic Phenomena of English.* Chicago: University of Chicago Press.

Mill, John Stuart. 1991. “Of Names and Propositions”. In Jay Garfield and

Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.

Montague, Richard. 1974. “The Proper Treatment of Quantification in Ordinary English”. In *Formal Philosophy.* Edited by Richard Thomason. New Haven: Yale University Press.

Neale, Stephen. 1990. *Descriptions.* Cambridge: The M.I.T. Press.
-- 1995. “The Philosophical Significance of Church’s Slingshot.” *Mind*, New Series, Vol. 104, No. 416 (Oct., 1995), 761-825.
--2001. *Facing the Facts.* Oxford: Clarendon Press.

Olson, Ken. 1987. *An Essay on Facts.* Palo Alto: CSLI Publications.

Quine, Willard van Orman. 1953. “Reference and Modality.” In *From a Logical Point of View.* Cambridge: Harvard University Press.
--1960. *Word and Object.* Cambridge: The M.I.T. Press.
-- 1966a. “Three Grades of Modal Involvement.” In *The Ways of Paradox.* Cambridge: Harvard University Press.
--1966b. “Vagaries of Definition”. In *The Ways of Paradox.* Cambridge: Harvard University Press.
-- 1966c. “Quantifiers and Propositional Attitudes”. In *The Ways of Paradox.* Cambridge: Harvard University Press.

Robinson, Daniel Nicholas. 2002. *Praise and Blame.* Princeton: Princeton

University Press.

Russell, Bertrand. 1903. *Principles of Mathematics.* Cambridge: Cambridge University Press.
--1990/1905. "On Denoting". In A.P. Martinich (ed) *The Philosophy of Language* Oxford: Oxford University Press.
-- 1920. *Introduction to Mathematical Philosophy.* London: George Allen Unwin.

Salmon, Nathan. 1986. *Frege's Puzzle.* Cambridge: M.I.T. Press.

Stalnaker, Robert. 1991. "Propositions". In Jay Garfield and Murray Kiteley (eds.) *Meaning and Truth.* St. Paul, Minnesota: Paragon House.

Strawson, Peter. 1990/1950. 'On Referring' in A.P. Martinich (ed) *The Philosophy of Language.* Oxford: Oxford University Press.

Tughendat. Ernst. 1970. "The Meaning of 'Bedeutung' in Frege". In *Analysis* (1970), pp. 177-189.

Weisler, Stephen. 1991. "An Overview of Montague Semantics." *Meaning and Truth.* 1991. Jay L. Garfield and Murray Kitely (eds.) St. Paul, Minnesota: Paragon House

Whitehead, Alfred North and Russell, Bertrand. (1997/1912). *Principia Mathematica (to *56).* Cambridge: Cambridge University Press.

Slippery Slope Paradoxes

An object moving one mph is moving slowly. If an object moving n mph is moving slowly, then so is an object moving n+1 mph. Therefore, an object moving 186,000,000,000 mph is moving slowly. True premises, a false conclusion: a paradox.

It is easy to generate similar paradoxes *ad libidum.* A person who has one dollar is not wealthy. If a person with n dollars is not wealthy, neither is a person with n+1 dollars. So someone with a billion dollars is not wealthy. True premises, false conclusion.

Let us refer to paradoxes of this type as "slippery slope paradoxes" (or SSP's, for short). An enormous amount of work has been done on SSP's. [304] Virtually *all* of the literature tries to solve the problem by asking us to abandon some accepted principle of logic.[305] So, according to one view, these paradoxes show us that some statements are neither true nor false or (what is perhaps different) are such that "there is no fact as to whether" they are true or false. (For some n, a person with n dollars is neither wealthy nor not wealthy. In any case, "there is no fact" as to whether such a person is wealthy or non-wealthy.)

According to another view, SSP's shows us that *modus ponens* (if P and [if P then Q], then Q) is not, in general, a valid form of inference. This, of course, allows one *both* to hold onto the plausible premises of an SSP *and* to reject the implausible conclusion.

Obviously we should tinker with the laws of logic as little as possible. In any case, a way of dealing with SSP's that *does not* reject modus ponens or bivalence is preferable (ceteris paribus) to one that does reject them.

Fortunately, the existence of SSP's in no way requires an alteration of our logic. We need only remember the obvious fact that, when people communicate with each other, various unstated and shared understandings are at work; and the sentences we use don't always make it clear what these understandings are.

Consider the sentence

(a1) "A person with 10,000 dollars is not wealthy."

Is (a1) true or false? That depends on the context. If you uttered (a1) in a country where the average person had four dollars to his name, you would be making a true statement. If you uttered (a1) in a country where the average person is worth hundreds of thousands of dollars, you would be making a false statement. So whether an occurrence of (a1) is true or false depends on context.

This suggests that when you use a sentence like (a1), some *benchmark* or *reference class* is taken for granted. And this benchmark or reference class will typically be the one that context most readily suggests. Suppose you are surrounded by billionaires, and you have just been talking about aspects of the tax-code that relate to billionaires. In that case, if you uttered (a1), people would rightly disagree. For in such a context, an utterance of (a1) would bear

the message:

(a2) Relative to people who have over a billion dollars, a person with 10,000 dollars is wealthy.

By contrast, if you were surrounded by people who had less than four dollars, and you uttered (a1), people would rightly agree with you. For in such a context, an utterance of (a1) would bear the message:

(a3) Relative to people with less than four dollars, a person with 10,000 is wealthy.

And, of course, (a3) is correct.

The main point here is that an occurrence of the predicate "is wealthy" involves some kind of comparison, often (but not always) an implicit one. An occurrence of "x is wealthy" bears the message "x is wealthy *compared* to__". And depending on how the "__" is filled out, the result is either true or false.

This point of view is not hard to substantiate. If you say "a person with 10,000 is wealthy", the person you're talking to can always ask: "Wealthy compared to what? What benchmark do you have in mind?" Suppose you counter-respond by saying "such a person is *just* wealthy; he isn't wealthy relative to this or that class of people, or relative to this billionaire or that indigent." Your counter-response is non-sensical. It is always reasonable to ask "wealthy compared to what?" Of course, in some cases, as we saw, context will supply the answer. But the question is always significant.

So any significant use of "wealthy" is really elliptical for "wealthy relative to__" or "wealthy for ___", where the "__" is filled out by expressions like

“Bill Gates” or “the average Swede”. Thus, any significant occurrence of a sentence of the form “a person with n dollars is wealthy” will be elliptical for “a person with n dollars is wealthy relative to__”. The same, therefore, is true of any sentence of the form “a person with n dollars is *not* wealthy.” (For this sentence is interchangeable with “it is not the case that: a person with n dollars is wealthy.”)

Once this fact about the predicate “wealthy” is recognized, it is easy to identify the fallacy embodied in SSP’s that use that predicate. A significant occurrence of “a person with one dollar is not wealthy” must be elliptical for a proposition of the form “a person with one dollar is not wealthy *relative to*__” If we fill out the “__” with “the average inmate of a POW camp”, what is conveyed will be false. If it is filled out with “the average citizen of Britain” or “the average accountant”, the message conveyed will be true. So if the sentence “a person with one dollar is not wealthy” is used in an argument, we must remember that, if it is being used significantly, it is elliptical for some comparative statement (e.g. *a person with one dollar is not wealthy compared to the average person* not *in a POW camp.*)

In a valid argument, different occurrences of a given term must have the same meaning.[306] The argument “JFK was a person; JFK was directed by Oliver Stone; therefore, some person was directed by Oliver Stone” looks just like a valid argument. But it isn’t one, because one occurrence of “JFK” refers to a person, while the other refers to a movie. So, in the argument given in the second paragraph of this paper, all the occurrences of the expression “not wealthy” must bear the same sense.

As we saw, that sense, when stated non-elliptically, will be given by some expression of the form “not wealthy *relative to*__” Of course, we must fill out the “__” in some way or other. (A proposition does not have blanks in it.) But – to close the argument – no matter how we fill out the “__”, the result is

an argument that has at least one *false* premise, and thus doesn't even appear to go through.

An example will help. Suppose we fill out the "__" with the expression "the average American accountant." In that case, the argument in question becomes:

(P1) A person who has one dollar is not wealthy relative to the average American accountant. (P2) If a person with n dollars is not wealthy relative to the average American accountant , neither is a person with n+1 dollars. (C) So someone with a billion dollars is not wealthy relative to the average American accountant.

P1 is true. But P2 is simply false. Suppose that x is the amount of money had by the average American accountant – in other words, x is the number that results when you add up to the number of dollars collectively had by American accountants and divide by the number of such accountants. x, of course, will be some specific dollar amount – let's say 27,000 dollars. In that case, P2 simply becomes *false*. For there *are* many numbers n such that, a person with n dollars is not wealthy relative to the average American accountant, and such that a person with n+1 dollars *is* wealthy by that standard: among such numbers are 27,000, 26,999.4, and 26,999.5. If we take the average American accountant as a kind of standard, and the average American accountant has 27,000 dollars, then technically someone who has 27,001 dollars is wealthy *relative* to the average American accountant. So the bold-faced argument simply contains a false premise, and thus doesn't even have the appearance of validity. So we dissolved the SSP without jettisoning any law of classical logic and without countenancing the idea that some statements are neither true nor false or (what may be different) are such that

there is no fact as to whether they are true or false.

Exactly similar considerations dissolve any SSP. Let us focus on the SSP given in the first paragraph of this paper. Suppose you say (*) “an object 35 mph is not moving quickly”, and your interlocutor responds by saying “quickly compared to *what*? The average glacier? The average electron? The average comet?” You could not legitimately counter-respond by saying: “an object moving 35 mph is *just* not moving quickly – no comparison or benchmark is intended.” Such a response would be meaningless. So, if you are speaking meaningfully, your utterance of (*) must mean something like “an object moving 25 is moving quickly relative to the average glacier” or “an object moving 25 mph is moving quickly compared to the average high-school sprinter at his top running speed.” So, in the argument in question, the predicate “quickly” must be seen as elliptical for some expression of the form “quickly relative to__”

But no matter how we fill out the “__”, the result is an argument that is *patently* invalid. Suppose we fill out the “__” with the expression “the top running speed of the average racing horse.” In that case, the argument in question becomes:

(P1) An object moving one mph is moving slowly relative to the top running speed of the average racing horse. (P2) If an object moving n mph is moving slowly the top running speed of the average racing horse, then so is an object moving n+1 mph. (C) So an object moving 186,000,000,000 mph is moving slowly relative to the top running speed of the average racing horse.

P2 is simply false. Suppose we take the top running speeds of all the race-

horses in existence, and divide that number by the number of race-horses. Suppose that 37 is that number. In that case, P2 is simply false. For there *are* many numbers n such that an object moving n mph is not moving quickly relative to the speed 37 mph, and such that an object moving n+1 mph *is* moving quickly relative to that speed: examples of such numbers are 36.8, 36.9, and 37. So the argument is simply a failure; and we showed that without rejecting modus ponens, or bivalence, or the law of excluded middle.

It is, I think, fairly clear that these considerations apply to any SSP one might think of. Given any predicate associated with such a paradox – "fat", "wealthy", "slow", "smart" – any significant use of that predicate must involve some kind of relativization to a reference class or benchmark. (One is smart relative to the average high-school student, not smart relative to the average Nobel Laureate.) So any occurrence of such a predicate has (if only implicitly) an empty slot, which must be filled out: "smart *relative to*__", "fat *relative to*__". When one fills out the predicate, and puts that newly filled out predicate back into the original SSP, the result is an argument with a false premise: an argument whose *in*validity can be easily established without jettisoning old friends like modus ponens and bivalence.

I will end by considering an objection to what we've said:

If someone has no money, then that person is absolutely non-wealthy: it is not he is non-wealthy relative to this or that standard. So, contrary to what you said, occurrences of "not wealthy" are not, at least in general, elliptical for "not wealthy relative to __" Analogous points hold of other predicates that generate PPS's.

Someone who has 0 dollars (or the equivalent in another currency) is

indeed non-wealthy in an absolute sense. But this means that such a person is non-wealthy relative to *every* standard – not that such a person is non-wealthy in a standard-less way. Of course, if a person is non-wealthy relative to *any* possible standard, then there is no need to cite any particular standard; it would be redundant to say "a person with 0 dollars is non-wealthy *compared to the average American accountant*." For it is patently obvious that relative to *any* standard, and therefore to that particular standard, such a person is non-wealthy. So being "absolutely" non-wealthy means being non-wealthy relative to *every* standard. Therefore the fact someone with 0 dollars is "absolutely" non-wealthy is in no way opposed to our contention that any meaningful use of "is not wealthy" must be capable of being fleshed out into a predicate of the form "not wealthy *relative to*__". Exactly similar remarks apply to all other predicates that generate SSP's. An object with an IQ of zero is, indeed, unintelligent in an "absolute" sense. But this means *not* that such an object is unintelligent in a standard-less way but, on the contrary, that such an object is unintelligent relative to *every* standard.

References

Vagueness**: A **Reader: Rosanna Keefe and Peter Smith (editors)
Cambridge, MA, and London: MIT Press, 1996

"Sorites", by Williamson, Timothy; and Sainsbury, R.M. In B. Hale and C. Wright (eds),
Blackwell Companion to the Philosophy of Language. Oxford: Blackwell 1999

The Sorites and Pseudo-monadic Predicates

Abstract: Some properties are ‘binary’ – one either has them or one doesn’t – while others can be had to different degrees. One cannot weigh 232lbs to different degrees – this is a ‘binary’ property. But one can be fat or intelligent to different degrees -- these are ‘degree’ properties. Degree-properties generate ‘Sorites’ type paradoxes: (i) a person with an IQ of 2 is not smart; (ii) if a person with an IQ of n is not smart, then a person with an IQ of n+1 is not smart; therefore (iii) a person with an IQ of 576 is not smart. Let us say that (i) is the ‘base’-clause in the argument just given and that (ii) is the ‘inductive’ clause. The way to block the inference to (iii) is, I submit, to see that sentences ascribing ‘degree’ properties are elliptical for *comparative* statements of a certain kind – statements that *relativizes* fatness or intelligence or tallness to some standard of comparison – and that this standard of comparison establishes a clear cut-off line between having and not having the offending property. Thus the inductive-clause is rendered false, and the paradox is not generated. This analysis of degree-properties is independently motivated. The concept of degree is bound up with that of measurement. And the concept of measurement is bound up with that of *comparison* (to some kind of standard). So x has phi to degree n iff x *compares* in a certain way to some standard.

I. If a person with an IQ of n is not smart, then a person with an IQ of n+1 won’t be smart. A person with an IQ of 4 is not smart. So, by induction, no matter how high an IQ a person has, that person is not smart. But obviously a person with an IQ of 356,567 is smart.

If a person is not heavy, adding one gram to his weight will not make him

be heavy. A person (of normal height) who weighs 100lbs is not heavy. So by induction, adding any finite number of grams to such a person's weight will not make him be heavy. So no one is heavy. But obviously a person (of normal height) who weighs 887lbs is heavy.

These are cases of the so-called 'Sorites' paradox. Any property that can be had in different degrees will generate this paradox. (People are intelligent to different degrees. People are heavy to different degrees.)

In this essay I will set forth what I think is a reasonable solution to this paradox. This solution is not in the literature.[307] But, so far as I can tell, it is no more noxious than the other solutions and should be considered, if only to be ruled out.

A disclaimer is in order. There are really many different Sorites-type paradoxes – we've already considered two different ones. What I will say probably does not constitute a solution to *every* paradox belonging to this class. But I think it may constitute a solution to some of them.

Of course, it would be better to provide a solution that applied to *all* Sorites-type paradoxes. But no such solution has ever been provided (at least none that is accepted by more than a tiny minority of the logico-philosophical community). And it is not immediately obvious that there should be a universal solution: I think that, beneath their superficial similarities, different Sorites-type paradoxes can have different logical structures and that, as a result, what works on one Sorites-type paradox may not work on another. In any case, a solution that applies to *some* Sorites-type paradoxes that *does* go through is preferable to one that applies to *all* Sorites-type paradoxes but *fails* to go through. And I'm confident that I will say goes through for at least some Sorites-type paradoxes

Some terminological points are in order. By an 'indeterminate' proposition I mean a proposition such that there is objectively no fact as to whether it is

true or not. By an ‘indeterminate’ concept I mean a concept such that, for some possible x, there is no fact as to whether x falls under that concept or not. And by ‘concept’ I don’t mean someone’s *grasp* of something; rather, I mean the meaning of a predicate --- the kind of thing that is the meaning of ‘fat’, ‘dirty’, ‘smart’.

I am leaving it open whether there really are indeterminate propositions or concepts. I am simply saying that *if* there is any proposition P such that there is objectively no fact as to whether P is true or not true, let us say that P is ‘indeterminate’; and that *if* there is any concept C such that, for some possible x, there is no fact as to whether x falls under C, let us say that C is ‘indeterminate’.

In section II, I give a rather dogmatic statement of my solution, i.e. I will provide a minimum of argumentation. In section III, I will provide an argument, and will try to discharge some of the more controversial assumptions underlying my solution. [308]

II. Sorites paradoxes are generated by arguments that have the following structure: a ‘base’ clause, an ‘inductive’ clause, and then a conclusion that follows from the base clauses but is known to be false. Consider the following paradigm Sorites-type argument:

(i) Someone with an IQ of 10 is not smart.

(ii) Giving an extra IQ point to someone who is not smart will not turn him into a smart person.

(iii) Therefore, by induction, someone with an IQ of (say) 12,242 is not smart.

(i) is the ‘base’ clause. (ii) is the ‘inductive’ clause. And, of course, (iii) is

the false conclusion that follows from (i) and (ii).

Consider any property that can generate Sorites paradox: large, dirty, strong, smart, bald. As we've already noted, what all of these properties have in common is that they admit of comparatives. One thing can be larg*er* – smar*ter*, dirti*er*, bald*er* etc. – than another.

Not all properties admit of comparatives: some are 'absolute.' The property of weighing 200lbs doesn't really admit of a comparative. It would be meaningless to say 'George weighs 200lbs to a greater degree than Fred', unless this were simply a tortured way of saying that Fred simply *doesn't* weigh 200lbs while George *does* weigh 200lbs. In general, given a property P that doesn't admit of comparatives, it would be meaningless to say 'x has P to a greater degree than y'. Obviously a corresponding point holds of the predicate corresponding to property P. (It would be as meaningless to say 'The predicate *weighs 200 lbs* applies to Bob more than it is to Jerry' as it would be to say 'Bob weighs 200 lbs more than Jerry.')

I submit that, for any property P that *does* admit of comparatives, it would be *similarly meaningless* to say 'x has P' *simpliciter*. There must always be some kind of relativization: some kind of a comparison or (what is related) some kind of implicit reference to a benchmark or standard of some kind.

There is no such thing as the proposition *Alpha is between* simpliciter or *Alpha is north* simpliciter. There is the proposition *Alpha is between Cleveland and Pittsburg* and *Alpha is north of Toledo* or *Alpha is north of here* (where the 'here' corresponds to a certain place). *Alpha is between* is certainly proposition-*like*; but it doesn't qualify as a proposition; it is missing something. I submit that the proposition *John is smart* simpliciter and *John is tall* simpliciter no more exist than the proposition *Alpha is north of* simpliciter and *Alpha is between* simpliciter exist. I submit that *John is smart* simpliciter falls short of being a proposition: it is missing something – like

Alpha is north of. What *John is smart* simpliciter is missing – what it needs to become a proposition – is some kind of relativization or completion: *John is smart* compared to 99% of the population; *John is smart* enough to figure out that we've been stealing from him. The underlined parts provide the aforementioned necessary completions or relativizations: without some such completion, *John is smart* simpliciter is, I submit, no more a proposition than *Toledo is to the north of* simpliciter is a proposition.

Of course, the *sentences* 'John is smart' and 'John is tall' are often *used* significantly. But when (e.g.) 'John is smart' is used significantly, it is (I submit) really elliptical for some proposition that relativizes John's level of intelligence to some standard: *John is smarter than most people* or *John is smarter than everybody in this department* or *John is smart enough to figure out such and such.* When 'John is tall' is used significantly it is really elliptical for *John is taller than 99% of the population* or *John is tall enough for such and such*: there is some kind of relativization – some kind of comparison or reference to a benchmark.

To state this point in terms of predicates: whenever a degree predicate (e.g. 'is heavy', 'is smart') is significantly used, there must be some kind of relativization to a standard or reference point: for any degree predicate P, what can significantly be said is *x has P to a higher degree than y* or *x has P to a high enough degree for such and such* or *x has P to a higher degree than everyone else in the old Folks home* or *x has P to a higher degree than 99% of other objects falling in class C* (for some class C) and so on. What *cannot* significantly be said is *x has P* simpliciter. If P really is a degree predicate then there is no more a proposition *x has P* simpliciter than there is a proposition *x is between* simpliciter or *x is to the north of* simpliciter: the 'entities' (I cannot say 'proposition', for that is not the right word) *x is to the north of* simpliciter and *x is between* simpliciter and *x is fast* simpliciter are

all in the same category – the category, presumably, of faux-propositions—the category of proposition-like entities that are, because of some failing or incompleteness on their part, not quite propositions.

As against this, one might say that *Alpha is between* and *Beta is the north of* are propositions, albeit non-sensical or insignificant ones. Very well; for the sake of argument, let us suppose that to be true. In that case, my view amounts to this. The proposition *Fred is fast* simpliciter is an insignificant proposition: what *is* a significant proposition is some 'completion' or 'relativization' of that insignificant proposition – e.g. *Fred is fast* enough *for such and such* or *Fred is faster than y* or *Fred is faster than everyone else on the U.S. track team*. For the record, my own view is it is a contradiction in terms of to speak of 'insignificant' propositions; for, in my view, a proposition just *is* a significance. So the entity *Alpha is to the north of* is, in my view, not a proposition, but some kind of a non-propositional concatenation of concepts – some kind of 'faux' proposition or proposition-like entity. All I am saying is that *if* one insists in saying that *Alpha is to the north of* simpliciter and *Toledo is between* are propositions, then my view can be accommodate that easily with a little terminological adjusting.

Let us sum up what we've said so far. Whenever a degree-predicate (a predicate that admits of comparatives) is used, it must be 'relativized' in some way. John cannot *just* be fast; he must be *faster than y* or *fast enough for x* (where x is some task or operation) or *fast to degree n.* Where there is use of a degree-predicate, but there is no such completion, no such relativization, the use is spurious – meaningless. So there is, strictly speaking, no proposition *John is smart* simpliciter (or, if you prefer, there is no *significant* proposition *John is fast* simpliciter); there is only *John is smart*er *than y* or *John is smart enough to do x* and so on. (Of course, what we've said about predicates applies *mutatis mutandis* to properties.)

This point constitutes the first step in my attempted solution to the Sorites paradox. Let us now move on to the next step.

Consider a specific Sorites paradox.

(a) A person of normal height who weighs 1000 lbs is heavy.
(b) Removing one pound from a heavy person will not make her stop being heavy.
(c) By induction, for any n, removing n pounds from a 1000 pounder will not make her stop being heavy (so if a 1000 pounder of normal height loses 950lbs, she will not stop being heavy).

Once it is seen that a meaningful ascription of heaviness to a person involves some kind of a relativization or reference to some kind benchmark, it is clear that it is impossible to generate the paradox just cited.

Consider the base-clause: a person who weighs 1000 pounds is heavy. If what we've been saying is correct, then the base-clause is either meaningless *or* it is really elliptical for some statement like *a 1000 pounder (of normal height) is heavier than 99% of the population* or *a 1000 pounder is sufficiently heavy that* ceteris paribus *her weight will cause health problems:* some statement that 'relativizes' the ascription of heaviness – some statement to the effect that a 1000 pounder is heavi*er* than y or *too* heavy for x or *so heavy* that such and such is inevitable.

Once the base-clause is replaced with some kind of meaningful relativization, then the inductive-clause becomes simply false. Replace the base-clause, i.e. replace (a), with some kind of 'relativization'; e.g. replace it with 'a 1000 pounder is heavier than 99% of the population.' If we take that relativization as the base-clause, then obviously the inductive clause no longer holds true: for *there is* some number n, such that given a person who

weighs n lbs, if that person loses one lb, then that person will *cease* to weigh more than 99% of the population.

Suppose Sally weighs 1000lbs. Everyday she loses one pound. There *will* come a day when, if she loses just *one more pound*, she will *right then* cease to weigh more than 99% of the population. Since there is such a day, the inductive clause is no longer true, and the paradox is no longer generated. Equivalently, *there is* some number n, such that someone who originally weighed 1000lbs need lose exactly n+1 pounds to stop weighing more than 99% of the population. So the inductive clause fails when we replace (a) with a suitable relativization; and the paradox isn't generated.

Another example may help:

(A) An object is moving at 10,000,000 mph is moving quickly.
(B) If an object moving quickly, it will *still* be moving quickly if it slows down by $1/10{,}000^{th}$ of an mph.
(C) So, by induction, an object that is moving at 1/100,000 mph is moving quickly.

The way to resolve this is to see that it is meaningless to say *an object moving at 10,000,000 is moving quickly* simpliciter. Obviously the words 'an object moving at 10,000,000 mph is moving quickly' (and 'Secretariat is moving quickly' and 'the Russian army is moving quickly') can be used to make significant statements. But when they were so used, they are elliptical for statements that involve reference to some kind of a benchmark: *an object moving at 10,000,000 is moving faster than anything we'd encounter on Earth, an object moving at 10,000,000 is moving too fast for our satellites to intercept it without being damaged, an object moving at 10,000,000 is*

moving too fast to be anything humans could build.

If you say ‘an object moving at 10,000,000 is moving quickly’, I can reasonably ask: ‘Quickly compared to what? Quickly with respect to what standard?’ Oftentimes the answers to these questions will be clear from the context of your utterance. But surely the questions just mentioned are always legitimate ones. And this suggests that (A) is really elliptical for some statement of the form *an object that is moving at 10,000,000 mph is moving quickly relative to such and such standard.*

Once that standard is specified, the inductive clause fails. One conceivable relativization of the sentence ‘an object moving at 10,000,000 mph is moving quickly’ – i.e. a relativization of (A) - might be ‘an object moving at 10,000,000 mph is moving too fast for our rocket to catch up with it.’ But given that relativization, (B) ceases to true. Obviously there *is* some number n such that if an object is moving at n mph, then our rocket *won’t* be able to catch up with it, and such that if that object is moving at n-1 mph, then (ceteris paribus) our rocket *will* be able to catch up with it.

Let us sum up. A relativization *must* be assigned to the base clause of any Sorites-type argument. Sentences like ‘an object weighing 1000 lbs is heavy’, ‘an object moving at 10,000,000 is moving quickly’, *unless* elliptical for relativizations, are meaningless– as meaningless as ‘x is north of’ or ‘x is between’. And *once* a relativization is assigned to the base-clause, then the inductive clause fails: the assignment of a relativization to the base-clause creates a kind of threshold that nullifies the inductive clause. And, of course, if the inductive clause fails, then the Sorites argument is invalid – as it then involves a false premise – and is thus innocuous.

III. As I see it, two points must be substantiated if the argument just given is to be made cogent.

(α): It is meaningless to ascribe baldness, heaviness, intelligence, etc. *simpliciter* to x; that any ascription of intelligence, heaviness, etc. must be 'relativized', if only implicitly, to some standard. (β): Any such relativization falsifies the inductive clause of a Sorites arguments.

A Defense of (α)

First of all, *prima facie* (α) is surely not unreasonable. If someone says 'x is moving fast' or 'an object moving at 100 mph is moving fast', it is always *significant* to ask: 'Fast relative to what? Fast with respect to what standard? Fast enough for what purpose?' Any object is fast compared to some objects and/or slow relative to other objects. Fast for a person is slow for a cheetah; fast for a cheetah is slow for a comet, and fast for a comet is slow for a beam of light. Suppose Bob is running at 30 mph. In that case, it is surely appropriate to say 'Bob is running outrageously fast.' But given a comet that was moving 30 mph, it would surely be inappropriate to describe *it* as moving 'outrageously fast': no sane person would do so. (Indeed, given such a comet, it would be legitimate to describe it as 'outrageously slow.') This obviously suggests that, when we say things like 'Bob is running fast'', we have some standard or reference point in mind: what we are saying is *with respect to some standard, Bob moving fast, i.e. compared to some standard, Bob is moving fast.* Of course, to say *Bob is moving fast with respect to some standard* is really to make a comparative statement – one like *Bob is moving fast enough for such and such* or *Bob is moving as fast as such and such* or *Bob is moving faster than such and such.*

Consider *an object that is moving at n mph is moving fast* (for any choice of n that you please).[309] If I say this to you, you can legitimately ask me 'fast compared to what? Fast with respect to what standard?' Suppose

I counter-respond by saying: ‘such an object is *just* moving fast: it isn’t moving fast *with respect to this or that standard*.’ Surely you have a right to be dissatisfied with my reply; indeed, my reply is nearly enough meaningless. Surely you deserve a response like: *an object moving at n mph is fast enough to orbit the earth ten times in an hour* (for some choice of ‘n’[310]) or *an object moving at n mph is moving faster than 99% of U.F.O.’s seen in this region.* Even if I say ‘an object moving at 186,000/second is *just moving fast*’ – 186,000 mph being the fastest speed physically attainable by an object[311] – what I am saying is *still* deficient: surely what must be said is along the lines of ‘nothing can move fast*er* than 186,000 mph/second: so anything moving that fast is moving *as fast as possible* – is moving *at least as fast* as any other object’ So even if n is assigned the highest value nomically possible, the sentence *an object moving at n mph is moving fast* is, if significant, elliptical for some statement that involves reference to a standard or to the speed of some other object or objects.

Here is another way to substantiate (α). Let us begin with a point made a moment ago: Uncontroversially, in response to the sentence ‘x is moving fast’ or *an object moving at n mph is moving fast*, it is always *significant* to ask ‘fast relative to what?’ Since it is always *significant* to ask for a completion of a sentence like ‘x is moving fast’ or *an object moving at n mph is moving fast*, that seems to show that the sentence *demands* a completion.

Consider: I say ‘x is moving fast’ or ‘an object moving at 98 mph is moving fast’. You ask: ‘fast relative to what?’ I say: ‘relative to nothing: it’s *just* moving fast.’ Surely my response is not satisfactory; surely it *would* be a reasonable counter-response to say: ‘x cannot *just* be moving fast; it must be moving fast *with respect to some standard* – it must be moving fast *for* a

comet (i.e. it must be moving fast*er* than most comets) or it must be moving fast *enough* to get us to New York before 6:00 p.m.' It is hard to see how it could always be *significant* to ask for some such completion, some such relativization, unless at some level such a relativization were *de rigueur*. (Compare: if I say 'John has many ideas', you can legitimately ask 'ideas about what?' Obviously the question 'ideas about *what*?' is legitimate because ideas *must* be ideas *about* this or that – about quantum theory or string quartets. So it is *legitimate* to ask for a completion – to ask 'ideas *about what*?' – because it is *mandatory* that there be a completion. In any discourse, it would seem, if a completion is possible, it is mandatory.)

Let us consider a different way of substantiating (α). It is hard to deny – and it is widely acknowledged – that Sorites-type predicates are 'context-sensitive.' (By a 'Sorites-type' predicate, I mean one that can generate a Sorites-type paradox: e.g. 'heavy', 'dirty', 'fast'.) An operating room that could legitimately be described as 'dirty' or even 'outrageously filthy' would probably be objectively cleaner by orders of magnitude than a room in a fraternity house that could legitimately be described as 'immaculate'. This context-sensitivity is easily explained if we see statements like 'this room is filthy' and 'this room is immaculate' as elliptical for relativizations – if, in other words, we regard tokens of sentences like 'this room is dirty' and 'this room is clean' as bearing propositions of the form *this room is clean compared to other rooms in fraternity houses* and *this room is sufficiently dirty that, if we operate in it, we will infect the patient.*

When we say that sentences like 'this room is dirty' are context-sensitive, that presumably entails that the *standard* we have in mind varies with context. The *standard* we have in mind when we describe an operating room as 'dirty' is different from the *standard* we have in mind when we describe a fraternity house as 'immaculate' or as 'not dirty'. So 'this room is dirty' is

elliptical for *relative to such and such a standard, this room is dirty.* And this bears out our thesis that such sentences *are* (when significant) elliptical for relativizations.

It seems that *all* Sorites-type predicates are context-sensitive. Consider 'heavy'. Plainly a sentence of the form 'x is heavy' is context-sensitive, i.e. whether such a sentence expresses a truth or a falsity depends on the context. 'Jerry is heavy' is true (expresses a true proposition) if Jerry weighs 200lbs and Jerry wants to be a jockey; and it is false (expresses a false proposition) if Jerry weighs 200lbs and he wants to be a Sumo wrestler. What makes a given token of 'Jerry is heavy' true or false is not just Jerry's physiognomy – though, of course, that is *a* determinant – but also the standard that is implicitly operative: one standard is appropriate when we are discussing Jerry's future as a jockey, and another is appropriate if we are discussing Jerry's future as a Sumo wrestler. So 'Jerry is heavy' is really elliptical for some proposition of the form *relative to such and such standard, Jerry is heavy*: which is consistent with our thesis that such sentence(-tokens) are, when significant, elliptical for relativizations.

There is a more formal and arcane reason to see Sorites-type sentences as elliptical for relativizations. (By a 'Sorites-type' sentence I mean the kind of sentence that can be the base-clauses in Sorites-paradoxes.) Consider the sentences

(i) *an object moving at n mph is moving fast*

and

(ii) *x is moving more fast than y.*

Presumably ‘fast’ has the same meaning in *both* (i) and (ii); we are not dealing with the kind of ambiguity exemplified by ‘bank’. In any case, it would be counter-intuitive in the extreme to deny that ‘fast’ has quite the same meaning in both these sentences. In (ii) ‘fast’ is part of a two-place predicate (__*more fast than*__) Now there can be no doubt that ‘fast’ is being used in an orthodox and completely significant way in (ii). Now if, in (i), ‘fast’ corresponds to a *one-place* predicate, then obviously ‘fast’ does *not* have the same meaning in (i) that it has in (ii): nothing could be more plain than that a one-place predicate cannot have the same significance as a two-place predicate. So given that ‘fast’ has the same meaning in both (i) and (ii); given that ‘fast’ is being used in a completely orthodox and unquestionably significant way in (ii); and given that ‘fast’ is being used as part of a two-place comparative predicate in (ii); it follows that (assuming (i) is being used significantly) ‘fast’ be being used in a comparative way in (i) as well.

The foregoing considerations, though a bit rough and ready, surely give some credence to the idea that Sorites-type predicates are (when used significantly) implicit comparatives: a significant token of *an object moving at n mph is moving fast* means *an object moving at n mph is moving fast compared to y* or *an object moving at n mph is moving is fast enough for z.* (The same is presumably true *mutatis mutandis* of ‘smart’, ‘heavy’ etc.) Some kind of a comparison is always implicit. Thus we have, at least to some minimal degree, discharged assumption (A): the base clause of an Sorites-type argument must be elliptical for some kind of relativization or completion: for some proposition of the form *an object that has P to degree n has P to a high*er *degree than y* or *an object that has P to degree n has P to a high* enough *degree for such and such condition to be satisfied.*

A Defense of (β)

(β) is not, I think, so contentious. Consider any statement of the form *x has P to a greater/less degree than y* (and statement of resembling forms like *x has P to a greater degree than most y's/99% of y's/all y's/…*). For example, consider an arbitrary statement of the form 'x is moving more slowly than y'. Obviously it is *not* true that, for each n, if x is moving n mph and is moving more slowly than y, then if x were to move n+1 mph, x would *still* be moving more slowly than y. I.e. there is some n such that, if x were moving n mph and were moving more slowly than y, then if x were to start moving at n+1 mph, x would *ipso facto* be moving faster than y.

An example may help. Consider the statement *x is moving more slowly than 99% of comets in our solar system.* Obviously it is *not* the case that, for each n, if x is moving at n mph and x is also moving more slowly than 99% of the comets its solar system, then if n were to increase its speed by 1 mph, x would *still* be moving slower than 99% of the comets in the solar system. In other words, *there is* some n, such that if a comet moving n mph is moving slower than 99% of the comets in its solar system, then if x were to start moving at n+1 mph, it would thereby *cease* to be moving slower than 99% of the comets in its solar system.

The point is that the inductive-clause in Sorites argument won't go through once a 'relativization' is associated with the base-clause. Suppose x is moving more slowly than y/is dumber than y/is lighter than y/…In that case, there *will* be some number n such that, if x has n IQ points/weighs n lbs/is moving n mph/….x *will* be slower/dumber/lighter/ …than y and such that, if x has n+1 IQ points/moves at n+1 mph/weighs n+1 lbs/…then x *will* be smarter/heavier/faster than y.

And consider statements of the form *x has P to a sufficient degree to do z*:

e.g. *x is strong enough to lift z,* where z is a 300 lb weight. Clearly *there is* some number n such that, if x can lift exactly n lbs, then x cannot lift z, and such that if x can lift n+1 lbs, then (ceteris paribus) x *can* lift z. Surely the same applies to *all* statements of the form *x has P to a sufficiently high degree to do z.*

One last example: Consider the sentence *an object moving n mph is moving fast.* Suppose we are right to say that this sentence, if significant, must be elliptical for some completion – e.g. *an object moving at n mph is moving fast enough to get from Los Angeles to New York in under three hours.* Plainly here is some n such that, if x is moving at n mph, then x won't be moving fast enough to get from Los Angeles to New York in under three hours and such that, if x is moving at n+1 mph, then (ceteris paribus) it *will* get from Los Angeles to New York in under three hours. So the inductive clause of Sorites-arguments won't work on statements of the form *x has P to a great enough extent to satisfy condition C.*

I would like to tie up a loose end. If somebody has *no* money, then it certainly seems reasonable to say that she is non-wealthy *in an absolute sense.* The same is true *mutatis mutandis* of all other Sorites-type predicates. And this seems to controvert my point that 'wealthy' (and therefore 'non-wealthy'), 'tall', 'smart', 'moves quickly' etc. have only a *comparative* meaning.

Let us consider the predicate 'non-wealthy'. What we will say will apply to all other Sorites-type predicates.

We've seen some reason to believe that 'non-wealthy' is typically used to make a comparative statement (though the standard of comparison is often implicit). So if we say that 'non-wealthy' has a truly *absolute* and non-comparative meaning, then we are really saying that it has *two* meanings: we are really saying that 'non-wealthy' sometimes expresses a non-comparative

predicate and sometimes expresses a comparative predicate, and is thus ambiguous between two very different significances.

But it seems extremely arbitrary to say that ‘non-wealthy’ is sometimes an absolute (one-place) predicate and sometimes a comparative (two-place) predicate. Obviously ‘non-wealthy’ has the same meaning in ‘Bob is *absolutely* non-wealthy’ (he has *no* money) that it has in ‘Bob is comparatively non-wealthy’ (he has less money than others in a certain reference class). And there is no reason to believe that the term ‘non-wealthy’ corresponds to predicates with different numbers of argument places in these two sentences; it is simply not natural to suppose that such a dramatic shift in its behavior occurs.

The natural way to go is this. To say that Bob is *absolutely* ‘non-wealthy’ is really to say that *relative to any standard at all* he is non-wealthy. To say that a light-beam is ‘fast’ *simpliciter* is really to say that *relative to any (physically possible) standard*, a light beam is fast. If a function is true for all values of its variables – e.g. ‘x is green or not green’ – logicians will sometimes describe it as being simply ‘true’. But nothing containing a free-variable is true (as these logicians of course know). Similarly, no sentence of the form ‘x is P’, where P is a Sorites-type predicate, is *simply* or *absolutely* true. It is true relative to this or that standard. But obviously if it is true relative to *every* standard, then there is no point in any citing any *particular* standard; and this makes it seems as though *no* standard is at work – when, in fact, just the opposite is true.

I would like to end by reaffirming one point. Many sentences whose *grammatical* form is ‘x has phi’, where phi is a one-place predicate, have a *logical* form involving a many-place predicate (i.e. a predicate of two or more places). The obvious example has to do with predicate ‘is moving’. ‘x is moving’ is grammatical of the form ‘x has phi’, where phi is a one-place

predicate, (In other words, ‘is moving’ is grammatically one-place.) But, we now know, motion is necessarily relative: x cannot move *simpliciter*; x can only move *with respect to y*, where y is some object or framework.

But what is obviously the case about ‘is moving’ (and ‘moves’ and so on) is less obviously true of many other predicates. Grammatically, any statement of the form ‘x is money’ is of the form ‘x has phi’, where phi is a one-place predicate. But, in reality, a bit of paper is money only *relative* to some system of customs etc. Nothing is money *simpliciter*: if there were no people in existence, a dollar bill would just be a piece of paper. *Grammatically* the sentence “the sound “Socrates” is a meaningful sign’ has the form ‘a has phi’, where ‘phi’ is a one-place predicate. But the sound ‘Socrates’ is a sign only *relative* to people who speak a certain language or, in any case, who adhere to certain norms. *By* itself, that sound is not a sign at all. So, in terms of its *logical* form “the sound ‘Socrates’ is meaningful” is really elliptical for some complex relational statement – one of the form ‘the sound “Socrates” is meaningful *relative to a group of beings characterized by such and such practices…*’ or, in any case, by some statement *not* of the form ‘a has phi’. Arguably, even a statement like ‘that is a hammer’ is, in terms of its logical form, not really a statement of the form ‘a has phi.’ For nothing is a hammer *simpliciter*; something is a hammer only relative to a person, or group of people, with certain intentions, practices etc. So, at least arguably, ‘that is a hammer’ expresses what is in fact a very complex relational statement, even though, in terms of its surface structure, it is completely non-relational.

All I am trying to do is to chip away at the main resistance that people will to the analysis set forth in this paper. I said that any statement of the form ‘a has psi’, where psi is a sorites-type predicate, is in fact a comparative of some kind. The obvious problem is that such statements seem *not* to be comparatives at all. Comparative propositions express relations (‘x is taller

than y' expresses a relation between x and y), and 'x is tall' doesn't seem to express a relation at all. But this *appearance* is not necessarily definitive. For, in general, grammatical form is often not a good guide to logical form. More specifically, statements that seem to have the form 'a has phi' often express relations. And, as I think we've shown, there is some independent reason to think that, in fact, sentences like 'Bob is tall' – sentences that serve as the base-clauses of Sorites argument – are implicitly comparatives of a certain kind.

References

Cargile, James. 1999, 'The Sorites Paradox'. in *Vagueness: A Reader*, Rosanna Keefe and Peter Smith (eds.), Cambridge: The M.I.T. Press.

Dummett, Michael. 1978, 'Wang's Paradox', in *Truth and Other Enigmas*, Cambridge: Harvard University Press.

Fine, Kit. 1999, 'Vagueness, Truth and Logic', in *Vagueness: A Reader*, Rosanna Keefe and Peter Smith (eds.), Cambridge: The M.I.T. Press.

Parsons, Terence. 2000, *Indeterminate Identity: Metaphysics and Semantics*, Cambridge: The M.I.T. Press.

Sorensen, Roy. 1998, *Blindspots*, Oxford: Clarendon Press.

Tye, Michael. 1999, 'Sorites Paradoxes and the Semantics of Vagueness', in *Vagueness: A Reader*, Rosanna Keefe and Peter Smith (eds.), Cambridge: The M.I.T. Press.

Williamson, Timothy. 1990. *Identity and Discrimination*, Oxford: Basil Blackwell.
--1994. *Vagueness*, London: Routledge.
--1999a. 'Vagueness and Ignorance', in *Vagueness: A Reader*, Rosanna Keefe and Peter Smith (eds.), Cambridge: The M.I.T. Press.
--1999b. 'Vagueness, Truth and Logic'. In *Vagueness: A Reader*. Rosanna Keefe and Peter Smith (eds.). Cambridge: The M.I.T. Press.

Williamson, Timothy and Sainsbury. R.M. 1998, 'Sorites', in *A Companion to the Philosophy of Language A Companion to the Philosophy of Language*, Bob Hale and Crispin Wright (eds.), Oxford: Basil Blackwell.

Wright, Crispin. 1999. 'Further Reflections on the Sorites Paradox', in *Vagueness: A Reader*, Rosanna Keefe and Peter Smith (eds.), Cambridge: The M.I.T. Press.

Sainsbury, R.M. 1999. 'Concepts without Boundaries', in *Vagueness: A Reader*, Rosanna Keefe and Peter Smith (eds.), Cambridge: The M.I.T. Press.

The concept of a symbol and the vacuousness of the symbolic conception of thought

Abstract: According to many research programs in philosophy and psychology, we think *in* symbols. I argue that, on pain of a vicious regress,

these internal symbols cannot themselves be grasped through symbols. At the same time, anything that is functioning as a symbol is necessarily *grasped*. Thus the symbolic conception of thought (SCT) actually entails that we grasp some things non-symbolically, and is therefore self-undermining. SCT must *deny* that something must be grasped to function as a symbol. I show that there is no way to make this maneuver work. SCT consequently ends up being the trivial view that we grasp things or the false view that symbols needn't be grasped.

I. According to a viewpoint that is currently very popular, thought is *symbolic*: representational content is accessed entirely through symbols. Let us refer to this viewpoint as the "symbolic conception of thought" (SCT). [312]

The symbols in question are not usually held to be those of some learned language like English or Albanian, but one belonging to some innate, internal code. And the act of accessing information through this code is held to occur at a *sub*-personal level. So this information-accessing is not merely unconscious, but *deeply* unconscious: psychoanalytic intervention will not make it become conscious the way such intervention would make your hostility towards your best friend become conscious.

This last point may require clarification. Surely *you* do not carry out the symbolic processes that, allegedly, underlie sense-perception and language-comprehension. After all, *you* do not generate your own sense-perceptions the way *you* compute the amount of money in your bank account. So those symbolic processes are not merely sub-conscious but sub-*personal*.

Whether SCT is correct is acutely important in this particular day and age.

These days many leading research programs in psychology operate on the assumption that the mind is a computer of sorts. Now *if* the mind is a computer, then presumably it accesses information through strings of code – through things that are, in effect, symbols of some sort. So if thought is not symbolic – if information is not typically accessed through symbols – then it seems to follow that we are not computers. And this, of course, would devastate the aforementioned research programs. [313]

The purpose of this essay is to show that, depending on how it is interpreted, SCT is either completely trivial or it is simply incoherent. More specifically, I will show that it coincides either with the platitude that some mental things are representational or with the absurdity that we grasp things which we do not grasp.

I must make it clear what I will and will not be attacking. I *do* believe that Chomsky, Marr, and other cognitive scientists are right to believe that language-comprehension, sense-perception, and other cognitive achievements derive from sub-personal information-processing. What I deny is that the information in question is typically accessed through *symbols.*

It cannot be denied that *some* information is accessed through symbols – or, at any rate, with the assistance of symbols. But, I will argue, it is not feasible to maintain that *all* information is accessed by way of symbols. The structure of the concept of a symbol does not permit this.

Traditionally, there have been two main lines of attack on SCT. First, this idea implies some kind of regress. To understand a symbol, one must know what meaning to assign it. And grasping a symbol does not *by itself* constitute knowing its meaning. (I can obviously grasp a symbol without knowing its meaning. That is why there is a need for translators, code-breakers, and so on.) So understanding a symbol involves grasping something *additional* to that symbol: it involves grasping the meaning of that symbol. But, according

to SCT, all information, and thus all meanings, are given through symbols. So, if SCT is right, then to understand a given symbol S_1, one must grasp some other symbol S_2, where S_2 gives the meaning of S_1. But if S_2 is to give one the meaning of S_1, then presumably one must understand S_2 – one must grasp *its* meaning. By logic exactly analogous to that just given, one must grasp that meaning through some third symbol S_3. But then one must grasp some *fourth* symbol S_4 to understand S_3 – and so on *ad infinitum*. So we are launched on an infinite regress.

The obvious way to stop this regress is to say that, at some point, one grasps a meaning without grasping a symbol which expresses that meaning. But, of course, that option is not countenanced by the symbolic conception of thought. So to avoid a vicious regress, we must throw out SCT.[314]

The typical SCT *counter*-response is to say that we *don't* know the meanings of the symbols which mediate our thoughts – we are simply built to *conform* to them.[315] The problem with this, of course, is that symbols must be *understood* if they are to function *as* symbols.[316] If we *don't* understand some putative symbol S, then *ipso facto* S is not a symbol *for us* – though it might be for someone else. Proponents of SCT must contest (and do contest) the inference from "S is not to understood by so and so" to "S is not a symbol for so and so".[317] We will spend a great deal of time showing that this inference is valid and that, consequently, the just mentioned SCT-ploy involves a non sequitur.

The other main line of attack on the symbolic conception of thought is this. The meaningfulness of symbols is necessarily derivative of the existence of thought. "Dog" means *dog* because of what people mean by "dog" – because, in effect, of the thoughts behind usage of "dog". The same is

presumably true of all symbols. So the existence of meaningful symbols presupposes the existence of thought. Therefore thought cannot itself, in general, consist in operations with symbols.[318]

The criticism I will put forth here is similar both of these other criticisms. But it doesn't presuppose the correctness of either. And, as we will see, it is therefore invulnerable to the counter-criticisms that have been leveled at them. In fact, it will itself be a source of new and non-question-begging responses to those counter-criticisms.

II. For the sake of argument, let us suppose that SCT is correct. In that case, you grasp the meaning *dog* (which, let us suppose, is the property of being a dog: the property of dogginess) through some mental symbol. Let DOG be this mental symbol. It must be asked: Do you grasp DOG or do you not grasp this symbol? If we say that you grasp it, then we must say that you grasp it through a symbol (for SCT just *is* the thesis that everything that is grasped is grasped through a symbol). Let DOG_1 be that symbol. Now do you grasp DOG_1? If we say that you do, then – so long as we adhere to SCT – we must say that we grasp it through some third symbol – call it DOG_2. It is obvious that there is a vicious and infinite regress if, for each n, we maintain that you grasp DOG_n through some symbol DOG_{n+1}: we end up saying that you must grasp infinitely many symbols to gasp a single one, which is implausible and probably incoherent.

But remember that, for SCT, everything that is grasped must be grasped through a symbol: even symbols themselves. We just saw that we incur a vicious regress if we say that each symbol is grasped through a symbol. The only way to get SCT to avoid being committed to this vicious regress is to say that we do not, at least not in general, grasp the symbols through which we grasp representational content.

But symbols are necessarily things that are *grasped.* Something can function as a symbol only if it is grasped. Consider any paradigm case of a symbol, say the English word "dog." Obviously you grasp this thing. And that fact, surely, is essentially connected with the fact that it functions as a symbol for you: if you didn't grasp it, it wouldn't be a symbol for you at all.

This point obviously generalizes to all paradigmatic symbols. Any given word of a language is a symbol for you because, inter alia, you grasp it. Snow, at least when referenced in poetry, symbolizes death to you because, inter alia, you grasp snow (you can see it and think about it).

Now we can close the argument. SCT says that everything is grasped through a symbol. We've seen that, in general, symbols cannot themselves be grasped through symbols. They must, at least some times, be grasped directly: otherwise a vicious regress is generated. But for something to function as a symbol, it must be grasped. So *if,* as SCT maintains, there are mental symbols, then these things must be grasped. At the same time, they are grasped directly, and not through symbols. So the symbolic conception of thought defeats itself.

There is a related argument for essentially this point. Suppose that DOG is one of the ungrasped mental symbols required by SCT. In virtue of *havin*g DOG, we grasp the property of dogginess. But it is not in virtue of *grasping* DOG that we grasp dogginess. For DOG is not grasped at all. So we don't grasp dogginess by way of *grasping* DOG or, more generally, by way of grasping *any* intermediary. *A fortiori* we don't grasp dogginess through any *symbolic* intermediary. We grasp it *directly.* So DOG ends up being a direct and thus non-symbolic grasp of dogginess; it ends being merely a garden variety concept. This point applies *mutatis mutandis* to all of the mental words posited by SCT: they all end up being garden-variety concepts -- direct, non-symbolic apprehensions.

So SCT ends up being the trivial view that we have concepts of things. To avoid this fate, SCT must say that we *do* grasp the mental symbols which it posits. (For if SCT says this, then it says we grasp *dogginess* by grasping a symbol of dogginess, and thus by way of a symbolic intermediary, as opposed to directly.) But in that case, as we've seen, it must then hold that we grasp those symbols *non*-symbolically, in which case SCT is false.

Summary

SCT just is the position that, whatever we grasp, we grasp it through a symbol. But we've seen that, if there are mental symbols, we don't grasp *them* through symbols. At the same time, something must be grasped to function as a symbol. So we must grasp those symbols. So SCT requires that there be things that we grasp non-symbolically. Thus SCT is inconsistent with itself and is therefore false.

II. The proponent of SCT will respond as follows (henceforth, all statements in single space are to those of an imaginary interlocutor):

You have insisted that, for something to function as a symbol, it must be *grasped.* You have merely taken this for granted. But I see no reason to believe this assumption true, and proponents of SCT don't typically hold it. But unless that assumption is true, your whole argument collapses.

A little background will help make it clear why something can function as a symbol only if it is grasped. First of all, not *all* representational entities have to be grasped do their job. You have a concept of the number five. But you do not have to *grasp* that concept for it to function – for it to enable you

to think about the number five. To think about the number five, it isn't necessary that your own *concept* of the number five be an object of your thought. (You *can* obviously think about your concept of the number five. But thinking about the number five does not necessarily involve your doing so.) So a *concept* doesn't have to be grasped to serve its function. (A concept is necessarily a grasp*ing*, but it is not necessarily something that is itself grasped.)

This point about concepts applies to many other kinds of representational entities. For a perception of a tree to pick out that tree, it isn't necessary that the subject make that very perception into an *object* of his thought. When you look at a tree, you are not necessarily thinking about that same looking; the object of your thought is the tree, not the fact that you are seeing the tree. So perceptions don't have to be grasped – don't have to be objects of thought – in order for them to do their job.

But can something function as a *symbol* without being grasped? The answer, I will argue, is 'no.' This becomes clear when we consider a related question: What is it that makes something qualify as a symbol *as opposed to* a non-symbolic representational entity?

A position that is *not* available to the proponent of SCT is this:

(S) Anything that represents something else -- anything that has representational content – is *ipso facto* a symbol. Concepts are graspings of things; they represent things. So *ipso facto* concepts are symbols.[319]

S is not available to the proponent of SCT because it utterly trivializes his position: it turns his position into the platitude that we grasp things. S says that any grasp of anything is *ipso facto* a symbol. Given this, the statement that we think in symbols becomes completely innocuous. If anything that is a

grasp of something else is *ipso facto* a symbol, then the statement "we grasp things in symbols" becomes a pleonasm; it then amounts to "we grasp things by means of things that are graspings", which is trivial, and says nothing more than that we grasp things. But SCT is not, surely, the innocuous platitude that we grasp things; everybody admits we grasp things, but presumably not everyone is a proponent of SCT.

This can be put another way. If *everything* representational is *ipso facto* a symbol, then SCT amounts to the uncontroversial and innocuous idea that some mental things are representational. Consequently, if SCT is to be non-trivial, then not *all* representations can be symbols.

But then we must ask: *Which* representational entities will a proponent of SCT be willing to describe as *non*-symbols? Obviously he cannot say that any representational *mental* thing is a non-symbol; for then he would be saying that the mind can, at least in some cases, access information *non-symbolically* – which is exactly the negation of SCT.

So the proponent of SCT must say that some non-mental representational things are non-symbols. But this is as unacceptable as the last option. First of all, if something is *intrinsically* or *non-derivatively* representational – that is, if it is representational independently of anyone's *taking* it as a representation – then it very much seems that it is *ipso facto* mental. Surely being *intrinsically* representational is as good a sufficient condition[320] as any for being mental. So the proponent of SCT is in the following predicament. First of all, he must find something representational that is *not* a symbol – if he cannot do this, then his position is totally trivial. Further, he must find an example of such a thing among representations that are *not* mental.

But this cannot be done. For any representation that is *not* mental is, quite plainly, only derivatively representational. ("Socrates" means *Socrates*

only derivatively of human mental contents. If there were no minds, and no mental contents, the inscriptions "Socrates" would mean nothing.) Presumably, the same is true *mutatis mutandis* of any non-mental representation. So any such representation is a representation only because, in effect, *we treat it* or *take it* as a representational. And this seems to mean: any such representation is a representation only because it is a *symbol* of some kind: something that we take to stand to stand for something else.

Let us sum up. To make his position non-trivial, the proponent of SCT must identify something representational that is not a symbol. He cannot identify any representational *mental* thing that is a not a symbol; for his whole position just *is* that mental representations are all symbols. So he must identify something non-mental that is representational. But any such thing seems very much to qualify as a symbol. Anything *non-mental* that is representational is only derivatively representational: somebody *takes* it to stand for something else, and that is what makes it be representational. But if somebody *takes* x to stand for y, then x is *ipso facto* a *symbol* of y.

So SCT ends up being trivial – for *everything* representational, on that view, ends up being a symbol. Thus the concepts *representation* and *symbol* end up being the same concept. Which means: the view that the every mental representation is a symbol ends up being trivial – ends up being the trivial view that every mental representation is a representation.

A proponent of SCT might try to extricate himself from this by weakening his position. He might say this:

Some cognitive activity is symbolic, but not all. So not *every* thought – and thus not *every* mental representation – is *ipso facto* a symbol. Thus the concept *symbol* becomes different from the concept *representation.* And SCT is there by saved from triviality.

The problem here is that this view is uncontroversial and trivial. Obviously we use symbols in our thinking all the time. No one denies this.

But the proponent of SCT has a response:

(OA[321]) It is uncontroversial that some mental activity is symbolic. But what *is* controversial and substantial is the view that in *some* cases our *only* access to some concept or, more generally, to some kind of representational content is through some kind of mental symbol. It is non-trivial to maintain that we can *only* grasp the square root of 2 or the concept of justice through a mental symbol. SCT – in the weakened version in question – maintains just this. So *in some cases* our grasp of something is entirely symbolic – *is* a mental symbol. And thesis is not trivial.

(OA) is indeed non-trivial. But, I submit, that view is wrong. My argument is this. The one thing which distinguishes symbolic representations from non-symbolic representations is this: symbols represent what they do because of how we *take* them; and this is not the case with non-symbolic representations. Now in order to *take* x to mean y, one must be able to grasp y *independently* of x. I couldn't possibly *take* "snow" to mean *snow* unless I had some way of thinking about snow that did not involve the symbol "snow." I couldn't take the sentence "there are more reals than rationals" to mean *there are more reals than rationals* unless I had some way of grasping proposition that *didn't* involve grasping the corresponding sentence. To take x to mean y is to assign y to x; and to do that, one has to have some grasp of y that doesn't involve x. So everything that is grasped symbolically is *ipso facto* also grasped *non-symbolically*. Thus even OA – which is the weakest imaginable version of SCT – isn't viable.

As I see it, there is only one controversial step in the preceding argument, namely my assumption that:

(*) The *one thing* which distinguishes symbolic representations from non-symbolic representations is that the former represent what they do because of how they are *taken*, whereas this is not true of the latter.

If our argument against OA has an Achilles heel, it is (*). So if OA is to work – and, therefore, if *any* version of SCT is to work (for OA is the weakest version of SCT) – then (*) must be wrong. Thus the proponent of SCT must show that what makes x be a symbol of y is, or at least can be, something *other* than x's being *taken* to represent y.

But this cannot be done. To see this, let us consider some mental representation that is *not* a symbol and use that as a kind of benchmark, so as to show, by way of contrast, what it is that makes something qualify as a specifically *symbolic* representation. This will make it easy to expose the incoherence in OA and therefore in SCT.

First of all, what is an example of a non-symbolic mental representation? It seems to me that if *anything* is representational but not symbolic – and we've seen that there must be such things if SCT is to have any content at all -- then *sense-perceptions* are in this category. In sense-perceiving the desk, you are just seeing the desk; you are not grasping the desk by way of grasping something other than the desk. So it is hard to see how any symbolism could be involved. In any case, it seems safe to say that if *anything* representational is non-symbolic, then sense-perceptions are. [322]

Also, even if sense-perceptions turned out to be symbols, that wouldn't affect that validity of the argument we are about to give. That argument would go through if we substituted talk of sense-perceptions with talk of

anything else that is a non-symbolic representation.[323]

So to discover what distinguishes symbols from non-symbolic representations, let us compare the word “Socrates”, which is unquestionably a symbol, with sense-perceptions, which we have agreed, at least provisionally, are non-symbolic representations.

Of course, the *obvious* difference is “Socrates”, on the one hand, and sense-perceptions, on the other is this. “Socrates” means what it does because of what we *take* it to mean. But, as we’ve seen, the proponent of SCT must say that what makes “Socrates” be a symbol is something *other* than the fact that people take it in a certain way.

But what else is there about “Socrates” that distinguishes it from things are representational but not symbolic? To be sure, there are other differences. For one thing, “Socrates” is in some sense a *public* entity, while sense-perceptions seem to be, in a fairly clear sense, private. “Socrates” is public currency; it is used by many different people. But my sense-perception of Socrates is not public currency in this sense.

But this move is a failure. First of all, there clearly are private symbols – symbols that only one person can use. When doing arithmetic, I can use images of grains of rice as symbols of numbers; and surely these symbols are private.[324] So what makes something be a symbolic as opposed to a non-symbolic representation is *not* that it is public.

This is not to mention the mental symbols discussed by psychoanalysis. My dream-image of Satan symbolizes my department head. Of course, some deny that there is any unconscious mental activity or therefore any dream-symbolism. But to say this is surely to take a very hard line.

Second, even if – what is clearly not the case – the property of being public *is* what distinguished symbols from non-symbols, that wouldn’t help

SCT. For the symbols posited by the proponent of SCT are every bit as private as any sense-perception.

Of course, there are yet other differences between “Socrates” and non-symbolic representational things. The symbol “Socrates” is realized by *non-mental* events and structures – by sounds and ink-marks. So this is certainly *a* difference between “Socrates”, on the one hand, and any sense-perception, on the other.

But this move is a failure. (Actually it is hardly different from the suggestion that symbols differ from non-symbolic representations in that they are public.) Some symbols obviously *are* composed of “mental” materials. As we saw a moment ago, mental images can serve as symbols, and mental images are obviously composed of mental entities.

Second, even if the property of being composed of non-mental materials *were* what distinguished symbolic from non-symbolic representations, this obviously wouldn’t help SCT at all. For clearly the symbols posited by SCT are *not* realized by non-mental things, i.e. they *are* realized by mental things. Those symbols, if they exist, are paradigmatically mental. They are thus unlike ink-marks or sound-tokens or anything else non-mental.

There is another difference between “Socrates” and a representational non-symbol like, for example, a sense-perception. Sense-perceptions are “analogue” representations, while words, and verbal representations generally, are “digital.” A sense-perception is not capable of decomposition in minimal content-bearing units: it has a kind of seamless or homogenous quality. By contrast, a verbal representation *is* capable of just such a decomposition. “Socrates punched Plato” is obviously capable of a decomposition into minimal meaning-bearing units (morphemes). So it is a digital representational. By contrast, a sense-perception (or series of sense-

perceptions) of Socrates punching Plato is not capable of decomposition into minimal, discrete parts; so it is analogue. So, to sum up, a possible difference between “Socrates” and representational non-symbols is that the former is digital, while the latter are analogue.

But this move, like the others, is a failure. For some symbols are analogue representations. Maps, radar screens, the contents on one’s computer “desktop” – these are all plainly symbolic representations. A radar screen with a blip on it symbolizes the presence of an airborne body in a certain location. But the radar screen is an analogue, not a digital representation. Surely what you see when you look at a “file” on your computer screen is not an actual file; it is an iconic representation, and thus a symbol, of an enormously complex network of electrical and mechanical interactions. But maps, radar screens, and desk-top icons are analogue, not digital, symbols. So being digital is not a property that separates symbols from non-symbols.

There may be other differences between “Socrates” and sense-perceptions. But – apart from the fact the fact that “Socrates” is derivatively representational, whereas sense-perceptions are non-derivatively representational -- it seems extremely unlikely that any of these differences is responsible for the fact that one of those things is a *symbol* while the other is not. The only reasonable conclusion is this:

(*) The essential difference between symbolic representations and non-symbolic representations is that the former mean what they do entirely because of what we take them to mean, while the latter mean what they do independently of what we take them to mean.

We’ve already seen how this fact – and we now know it to be a fact – is

deadly to SCT.

Of course, if x must be *taken* a certain way to represent y, then x must be *grasped.* (I cannot *take* x to mean y unless I grasp x.) So symbols are necessarily *grasped.* (This is what we set out to prove at the beginning of this section.) We have already seen how this fact devastates SCT.

A corollary of (*) is that the only significant thing which the so-called "symbols" posited by SCT have in common with clear cases of symbols – e.g. the word "Socrates" – is that both types of things are *representational.* Beyond that, there is no significant commonality. But this means that, in saying that we think in "symbols", the proponent of SCT is either wrong or he is using the word "symbol" to denote *anything* representational. But, as we've seen, if he is using the word "symbol" in such a broad sense, then his thesis that thought is "symbolic" becomes the trivial thesis that thought is representational.

I should address a possible objection:

There is one difference that you did *not* consider between symbolic representations and non-symbolic representations. One always *learns* what a symbol means. One has to *learn* that "Socrates" is the name of a certain philosopher. But one doesn't *learn* what a sense-perception represents. In virtue of having the sense-perception, you are *ipso facto* aware of what it represents.

Of course, if *you* come up with a word – if you decide to let "Blarb" be the name of your dog – then *you* don't have to learn what that word means. But other people still will. Other people must *learn* that your dog is named "Blarb."

So the difference between symbolic and non-symbolic representations is, ultimately, that we must *learn* what the former represent, but not what the

latter represent.

I actually think this point is correct. But, first, it is no help to SCT. Second, it ultimately coincides with our point that what symbolic representations mean what they do in virtue of what we take them to mean, whereas this is not true of non-symbolic representations. Let me now defend these assertions.

Obviously a proponent of SCT cannot say that we ever *learn* what mental symbols represent. For to *learn* that x means y, one must have some way of grasping x independently of y. If I had to *learn* that QUATERNION was my mental symbol for quaternions, that would mean that I had to learn that QUATERNION bore the meaning *quaternion*. But to do that, I'd have to have some grasp of the concept *quaternion* that didn't involve QUATERNION. So the objector's point doesn't help SCT or OA.

Let us now elucidate the second point. Everything the objector says is correct. I don't have to learn to what sense-perceptions represent: if I have a sense-perception, I *ipso facto* am aware of what it represents. Of course, I have to learn about the things given to me *through* sense-perception. A single sense-perception of water will not inform me that it consists of discrete particles. But if I have a perception of water, water is *ipso facto* the object of my awareness. I don't have to *learn* anything for the perception to make water be an object of awareness: the perception does that all by itself.

By contrast, one must always learn what a word means. By itself, the word "Socrates" doesn't put me in any kind of cognitive contact with the man. For "Socrates" to represent Socrates to me, I must first learn what it means. The same seems to be true of symbolic representations generally: we must always learn what a symbol represents, just as the objector states.

But that fact, I think, is derivative of the fact that symbols mean what we *take* them to mean. Nothing intrinsic to the verbiage "Socrates" makes it refer to a certain Greek philosopher. That same verbiage could represent a planet or a lake or nothing at all. The only reason it picks out a certain philosopher is because people *use* it or *take* it to do so. But a sense-perception of a certain tree will necessarily represent that tree: it doesn't matter how we "use" it or "take" it. (Indeed, these concepts have no application when it comes to sense-perception.) And that is obviously why we don't have to *learn* what the perception represents. The connection between the perception and the tree isn't derivative of anything; it is built right into the perception. So that connection comes *with* the perception. That is why you don't have to *learn* about that connection. Merely *having* the perception is enough.

So it is true that, ultimately, we must learn what symbols mean. But that derives from the fact that symbols mean what they do because they are only *derivatively* representational – because they mean nothing by themselves, but only what people *take* them to mean. So the objector is, in effect, conceding that what makes x be a symbol of y is that people take x to mean y. And, as we discussed earlier, this is just what a proponent of SCT and OA cannot concede. So we may safely conclude that what makes x be a symbol of y is specifically that people take x to mean y.

Before ending this section, I should address a possible objection to this last contention:

Clearly certain kinds of spots represent measles, and smokes represents fire. So smoke is a symbol of fire, and spots are a symbol of measles. But smoke is a symbol of fire *independently* of what anyone takes smoke to mean. Spots are a symbol of measles independently of what anyone takes them to mean. So you are wrong that what makes x be a symbol of y is that

people *take* x to mean y.

Smoke is not a *symbol* of fire. It is an *effect* of fire. The relation of smoke to fire is nothing like the relation of "fire" to fire. Of course, since smoke is an effect of fire, smoke "tells" us that there is (or was) fire. But that only means that we can infer that there is (or was) a fire from that the fact that there is smoke – just as we can infer that it recently rained from the fact that the streets are wet. But the wetness of the streets is not, by itself of rain; it is just an effect of rain; and it's relation to rain is not remotely comparable to the relation of the word "rain" to rain. Of course, people could *use* smoke as a symbol of fire: people can use anything to symbolize anything. We could, for instance, invent a language in which assertions about fire are made by using smoke. But in that case people would *take* smoke to mean *fire*, and that is why it would be a *symbol* of fire, and not a mere effect of it.

IV. As we discussed earlier, there are two traditional attacks on SCT: (i) it implies a certain kind of vicious regress. (ii) The intentionality of symbols is derivative of that of mental contents. So at least some representational mental entities must be non-symbols.

These criticisms are, it seems to me, very closely interconnected. In fact, they seem to be the same criticism. Both are to the following effect. In order for something to function as a symbol, it must be taken a certain way; therefore, on pain of some kind if incoherence, this taking cannot itself consist in the tokening of some symbol. But to avoid prejudging any questions, I will treat these criticisms as separate; the inner identity between them will naturally emerge in the following comments.

I want to show that what I have said in this paper is not vulnerable to

the traditional SCT counter-responses to (i) and (ii). First of all, what are these counter-responses?

The traditional counter-response to (ii) is simply to say that it is question-begging:

If in fact we think in symbols, then the intentionality of symbols is not always derivative of that of thought. There is independent reason to think that we think in symbols (e.g. we are computers, and computers think in code/symbols.) So maybe the old saw that symbolic intentionality is derivative isn't true.

Let us begin with this counter-response. We have seen that the *one thing* that distinguishes symbols from non-symbolic representational entities is specifically that symbols must be *grasped* and *taken a certain way* in order to function as symbols, while this is not true of non-symbolic representational entities (e.g. sense-perceptions). We have not merely taken this for granted; we have argued for it. Now given that symbols must grasped and taken a certain way to function as symbols, it is fairly clear that their intentionality *is* derivative of that of those graspings and takings, and thus *is* derivative of that of other mental entities. (If we say that those graspings and takings are themselves symbols, or symbol-tokenings, then a vicious regress is incurred.) So we have not merely *taken for granted* that symbolic intentionality is derivative of that of mental intentionality; we have actually discharged this intuition. So what we have said constitutes a non-question-begging response to the objector.

The traditional SCT counter-response to (i) is as follows. *If* we had to know the meanings of the mental symbols posited by SCT, then indeed there

would be a vicious regress; but we don't *know* the meanings of those symbols: "we are simply *built* to use it [those symbols] correctly", as William Lycan put it.[325]

Jerry Fodor (in private correspondence) put this point in a slightly different way:

> [Y]ou don't have to represent the rules of a language that you are able to use; all that's required is that you are so constructed as to conform to the rules. (Cf the way computers are `built to use' their machine language.) Of course, it's plausible that you DO internally represent the rules of (say) English; but that's not because English is a language; it's because it's a language that you LEARN; and it's quite plausible that to learn a language is to come to internally represent its syntax and semantics. (Presumably in "Mentalese").

The Fodor/Lycan position is incompatible with some basic truths concerning the nature of symbolism. To begin with, the expressions posited by SCT are not *public* – they are not like the expressions composing English. They are completely internal; they transmit information from one part of your mind to another.

Trivially, either the expressions of this internal language mean what they do because of how you (presumably at some sub-personal level) take them or they mean what do independently of how you take them. (Henceforth I will omit the qualification "at some sub-personal level.") Let us consider each alternative.

As we've seen, it is incoherent to suppose that a symbol means what it does independently of how anyone *takes* it. Now where the symbols mediating *your* thought are concerned, if *you* do not take those symbols in this or that way, then *nothing* takes those symbols in any way. Obviously,

another person cannot *take* the symbols internal to your mind in a certain way. In any case, another person's doing so is surely not what would typically give those symbols their meaning. So, with regard to the symbols that, according to SCT, mediate your thought-processes, those symbols mean what they do because of how *you* take them.

Given that these symbols mean what they do because *you* take them a certain way, it is obviously inconceivable that you should altogether fail to know what they mean. But once we concede that you know the meanings of these symbols, we *either* generate a vicious regress (*vide* the first page of this essay) *or* we concede that you grasp these meanings in some symbolic form.

But this argument involves a non-sequitur and thus doesn't go through. You say that the mental symbols posited by SCT mean what they do because of how the individual takes them. And you conclude from this that the individual must (at some level) know what these expressions mean. But this simply doesn't follow. Obviously the word "meticulous" means what it does because of how English speakers take this word. But that doesn't mean that each English speaker knows what this word means. Surely many English speakers don't know what it means.

English is a *public* language; the expressions belonging to it mean what they do because of how *many* people take them and use them. A given speaker can fail to know what a particular word means – e.g. "meticulous" – because there are many other people to give meaning to that word: for there are many English speakers. But the symbolic systems posited by SCT are private. If *you* (at some level) don't assign those expressions meaning, no one else will. So *you* and you alone give them meaning; and they remain

meaningful only as long as you continue to give them meaning. So, at some cognitive level, you must know their meanings.

Suppose that you were the *only* speaker of English. In that case, you really *couldn't* fail to know what a given expression of English meant. For if you forgot what a particular expression meant, it would simply drop out of English.

To take another analogy: suppose – what is, of course, the case – that English is spoken by many people, but that *all* of them stopped knowing what "meticulous" meant. In that case, "meticulous" would simply stop being a part of the English language. Since you are the sole guardian of the symbolic system in which, according to SCT, your thoughts are embodied, you cannot fail to know what the expressions of it mean, since you alone give them meaning.

You have said repeatedly that symbols mean what they only because we *take* them a certain way. But that is straightforwardly false. Consider the sentence: "if Brahms had been born in the 17th, as opposed to the 19th, century, he probably would have been the father of Baroque music." That sentence has a definite meaning, even though (in all likelihood) it has never been produced before, and *a fortiori* no one has ever *taken it* to have this or that meaning. Also, linguists are always coming with sentences (e.g. "the horse raced passed the barn fell") that have definite meanings, but that are almost always mis-interpreted. If sentences meant what they did entirely because of what people *took* them to meaning, then this wouldn't possible: systematic mis-interpretation would be correct interpretation. But systematic mis-interpretation is not correct interpretation, as linguists have shown.

There are, ultimately, two components to the semantic dimension of a language. There are the simple expressions composing it (e.g. “red”, “dog”), and there are also the recursive devices which assign meanings to complex expressions on the basis of the meanings of the simple ones.

Now it is undeniable that *simple* expressions mean what they do entirely because of how people take them. If everybody meant *cat* by “dog”, then “dog” would mean *cat*. It would be nonsense to maintain otherwise. Now it is also clear that the meaning of recursive devices is entirely determined by how people *take* them. A paradigm case of a recursive device is the adding of an “s” to a noun to pluralize it. Now suppose everyone *took* the adding of an “s” to a noun to denote (say) the number of legs had by typical specimens of things to which that noun applied. (This a perfectly legitimate example of a recursive rule.) In that case, “cats” would really mean *four*, “humans” would really mean *two*, and “snakes” would mean *zero*. So the recursion associated with adding an “s” to a noun would be completely different from what it is in actual English. Thus, recursive devices, no less than simple lexical items, mean what they do entirely in virtue of how people take them. Now it is true that *once* simple lexical items and recursive devices are assigned meanings by how people *take* them, then the meanings of complex expressions that people have never *taken* to mean anything are thereby fixed. But this poses no real threat to the thesis that symbols mean what they do because of how people *take* them. *Any* symbol, whether simple or complex, means what it does entirely because of how people take the simple symbols composing it.

This means that *any* symbol is either identical with, or composed of, symbols that have the meanings they do because of what people *take* them to mean. And this, of course, implies that any symbol is identical with, or composed of, symbols that are *grasped*. (An ungrasped symbol cannot be

taken to mean anything.) Once this is conceded, then to avoid a vicious regress, it must also be conceded that these symbols are grasped non-symbolically. And therewith the symbolic conception of thought falls through.

Let us consider another objection:

Your argument is basically this. SCT says that everything is grasped through a symbol. But symbols themselves cannot, in general, grasped through symbols; otherwise we have a vicious regress. At the same time, something must be grasped to be a symbol.

Though perhaps technically sound, this argument seems a bit petty. To survive it, SCT need undergo only the following, practically trivial amendment:

(**) *everything is grasped through symbols except mental symbols*. Those symbols are grasped directly (non-symbolically).

(**) is obviously a significant claim, and it is obviously in the spirit of SCT. You've done nothing to refute it.

The problem with (**) is that if any *one* thing can be grasped non-symbolically, then so can many other things. It isn't feasible to maintain that there is only *one thing* we can grasp non-symbolically. For if we a grant a single exception to SCT, countless others come in through the backdoor.

For the sake of argument, suppose that a person can grasp mental symbols non-symbolically. Obviously, one must be able to grasp different mental symbols – e.g. the mental symbols **DOG** and **CAT** – *as* being different; he must be able to register what distinguishes the one symbol from the other.

And the same is true, mutatis mutandis, for any other two mental symbols one might choose. Supposing that we do think in symbols – that we access propositions wholly through symbols -- there are indefinitely many different mental symbols (infinitely many, in fact: for there are infinitely many different thoughts a person can have); and you, the cognitive subject, must be able to register the differences between any two of them. To be able to recognize each of an infinitely large class of different symbols, you must have at your disposal an extremely extensive repertoire of concepts and of cognitive abilities. Things cannot *just* be grasped; they must be grasped *as having* these or those properties. You cannot *just* see Fido, or *just* think about Fido; you must see him, or think about him, *as* having a certain color, shape, dispositions, and so on. So you couldn't *just* grasp a mental symbol (we are still supposing, for the sake of argument, that there are such things as mental symbols); you would have to grasp it *as* having these or those properties. Further, you would have to grasp it in such a way that, on the basis of the information encoded in that grasping, you were able to distinguish that particular mental symbol from the infinitely many others which can you can grasp.

To see this last point, suppose for the sake of argument your grasp of the mental symbol **DOG** were *not* such that, on the basis of the information encoded in that grasp, you could distinguish it from the mental symbol **CAT**. In that case, these two symbols would be identical as far as you could tell. So, in effect, they *would* be identical, as the differences between them – not being registered by you – would be inert as far as your thinking was concerned. But then they couldn't function *as* different symbols. Thus, if the various symbols mediating your thought are to function *as* different symbols, then your grasp of any one of these things must encode enough information that, on the basis of that grasp, you can distinguish that symbol from the various symbols

which mediate your thought.

This means that any given grasp of a mental symbol must contain an enormous amount of differential information: your grasp of the mental symbol **DOG** must contain enough information if **DOG** is to continue to function as a distinct symbol, and to avoid collapsing into, say, your **CAT**-symbol or your **AARDVARK**-symbol. So your grasp of **DOG** must represent that symbol as being different from your **CAT**-symbol, your **AARDVARK**-symbol, your 1-symbol, your 2-symbol, your 998.87-symbol, and so on. So your grasp of **DOG** must present that symbol as having different *properties* from **CAT**, **AARDVARK**, **1**, **2**, **3**, and so on *ad infinitum.* This means that you have must some kind of cognitive access to an extremely wide array of properties (for obviously you cannot grasp something as having this or that property – as being, say, red or square – unless you have some kind of cognitive access to those properties). And, of course, *this* access cannot be mediated through *symbols* of those properties, since a grasp of those properties is a *pre*-requisite to grasping, and registering the distinguishing characteristics of, any symbol. So if we grasp mental symbols non-symbolically – and we've seen that we typically must, if indeed there are such things – that ability presupposes that we have non-symbolic cognitive access to a great many properties.

Further, *merely* having cognitive access to those properties will not be enough to register the distinguishing characteristics of the various symbols which, according to SCT, mediate our thought. We'd obviously have to *recognize* each mental symbol for what it was; and such a recognition, it is plausible to suppose, would itself involve various cognitive processes.

Surely mediating between your *seeing* Fido, on the one hand, and your *recognizing* Fido *in* what you see, on the other, are various cognitive processes. If you had a defective memory, you could *see* Fido – even have an

accurate, high-resolution of perception of him – and simply not *recognize* that the dog you're seeing is Fido. (Many people with Alzheimer's have perfectly good vision and hearing, but they may fail to *recognize* their own spouses. So they *see* their own spouses with perfect accuracy, but they don't recognize their spouses *in* what they see.) And presumably various mnemic and cognitive operations are implicated in the co-occurrence of that act of *recognition* with that act of perception. In any case, it is highly reasonable to suppose that such operations are involved.

And— supposing for the sake of argument that there are mental symbols -- it also seems highly reasonable to suppose that similar cognitive and mnemic operations are implicated in the co-occurrence of, on the one hand, your *grasping* (say) **DOG** and, on the other hand, your *recognizing* that that what you are thereby grasping is, in fact, that symbol. For just as you could have a perfectly accurate and detailed perception of Fido without *recognizing* Fido in what you are seeing, so presumably you could have a perfectly veridical and detailed grasp of **DOG** without *recognizing* that symbol in what you are grasping. So given that a great deal of cognitive activity is involved in the one case, it seems reasonable to assume that it is involved in the other. [326]

Of course, none of this implicated cognitive activity could be mediated through mental symbols; for it would all be *pre*-requisite to your grasping and also recognizing mental symbols.

To sum up, if we concede that we can recognize mental symbols non-symbolically, then we are by implication conceding that we have non-symbolic cognitive access to many other things (notably properties) and also that a great deal of cognitive activity is non-symbolic. So the objector's position s *not* one that a proponent of SCT can take. For that position doesn't merely circumscribe SCT: it undercuts it altogether.

In fact, once it is granted that we can grasp *symbols* non-symbolically, and once it is seen what a rich array of cognitive abilities is necessarily implicated in such grasping, it becomes clear that our thinking must be *essentially* non-symbolic. For it would surely be arbitrary to maintain that those multifarious cognitive abilities were deployed *only* in the grasping of symbols and in no other capacity. Surely those non-symbolic cognitive abilities permeate, if they don't actually constitute, our cognitive framework.

References

Fodor, Jerry

1975 *The Language of Thought* New York: Thomas Y. Crowell Co.

Horst, Stephen

1996 *Symbols, Computation, and Intentionality* Berkeley and Los

Angeles: University of California Press

Kripke, Saul

1972 *Naming and Necessity* Cambridge, MA: Harvard

Kuczynski, John-Michael

2002 Does the Idea of a “Language of Thought” Make Sense? *Communication & Cognition* 35 (3-4) 173-193.

2003 Are We Computers? A Wittgensteinian Approach *Existentia* XIII (3-4) 219-239.

Lycan, William G.

1984 *Logical Form in Natural Language* Cambridge, MA: The M.I.T. Press

Pylyshyn, Zenon W.

1984 *Computation and Cognition: Toward a Foundation for Cognitive Science* Cambridge MA.: The M.I.T. Press

Searle, John

1992 *The Rediscovery of the Mind* Cambridge MA.: The M.I.T. Press

1998 *Mind, Language and Society* New York: Basic Books

Wittgenstein, Ludwig

1974 *Philosophical Grammar* edited by Rush Rhees, translated by Anthony Kenny Berkeley and Los Angeles: The University of California Press

2002 *Philosophical Investigations* translated by G.E.M. Anscombe Oxford: Basil Blackwell (3rd Edition)

Grammatical inflexions in connection with the nature of Sentence-level Meaning

Consider the proposition

(*) *there is a cactus in front of my window*

A world where (*) is true is different from one where (*) is false. In the one

kind of world, there is an instance of the property of being a cactus in a certain place, at a certain time. In the other kind of world, there is no instance of that property. It thus seems that for (*) to be true is for various properties to be instantiated and that for (*) to be false is for those same properties *not* to be instantiated.

What is true of (*) is obviously true of other propositions. These reflections suggest a certain view as to what propositions are, and also as to what it is for a proposition to be true. A proposition is a collection of properties, and a proposition is true if the relevant properties are instantiated.

As it happens, (*) is false. But I can easily enough imagine what things would be like if it were true. I imagine a world where, at certain places and times, there are instances of certain properties. If (*) were true, I could easily enough imagine what things would be like if it were false: I would imagine a world where, at certain places and times, there were no instances of the properties just mentioned.

I thus propose that a proposition is a set of properties, and that for a proposition to be *true* is for all of those properties to be instantiated. Truth is identical with instantiatedness.

In this paper, I will develop the suggestions just put forth. There are atomic propositions and molecular propositions. The position in question can be made to apply to both.

§ There are two requirements that any adequate theory of propositions must satisfy. First, it must accommodate the fact that a propositions has a certain *composition* – or, as Katz put it, a certain *de*composition. Second – this is related – it must account for the *unity of the proposition*; it must explain how a proposition's various constituents can combine into something true or false, instead of remaining a truth-valueless set or heap.

The proposition *Socrates punched Plato* has a unique decomposition into minimal parts; and that proposition is obviously identical with a certain arrangement of those parts. The question is: *how* exactly are those parts arranged in that proposition? What conceivable arrangement of Plato, Socrates, and the relation of punching constitute a true proposition? The set (Plato, Socrates, the relation of punching) obviously isn't true or false. It doesn't matter whether we impose some kind of ordinal structure on that set. The *ordered* triple <Plato, the relation of punching, Socrates> isn't true or false. Neither is <the relation of punching <Plato, Socrates>>, or any other ordering of those three constituents. So even though it is clear that *Plato punched Socrates* is identical with *some* arrangement of those three things, it is far from clear what kind of arrangement is the appropriate one. In my view, it becomes quite clear what the right kind of arrangement is, so long as we pursue the suggestions made earlier regarding the nature of propositions.

(A) "Smith punches Jones"

is a sentence.

(B) "Smith, the relation of punching, Jones"

is not.

A sentence is not just a heap of referring terms. A corresponding point holds about propositions.

(Ap) *Smith punches Jones*

is a proposition.

(Bp) *Smith, the relation of punching, Jones*

is not. A proposition is not a heap of objects and properties. A proposition is a set of objects and properties that are *unified* in some way. But *how* must they be unified? In what does the "unity of the proposition" lie? This question has never been adequately answered. I believe that our analysis of propositions gives us an answer to this question.

§ Frege's answer was to say that, in a proposition, certain items are "unsaturated": they have holes in them. These holes enable the items in the proposition to link together in the right way. In (A), the expression "punches" corresponds to something with holes in it; these holes enable Smith and Jones to link up with that thing in the appropriate way. By contrast, in (B), the thing corresponding to "the relation of punching" doesn't have the right number of holes in it; so it doesn't appropriately link up with Bob and Smith.

Frege's "solution" is really just a way of labeling the problem, as Davidson said.[327] Obviously we aren't talking about physical holes. We are talking about holes in a metaphorical sense. What we want to know – what Frege doesn't tell us – is what this metaphor is a metaphor for.

The difference between (A) and (B) obviously has something to do with the fact that the first has grammar, whereas the second does not. In the first, there are inflections. (I will argue later some of these are not phonetically apparent.) In the second, nothing is inflected. The difference between a proposition and a non-propositional heap is somehow coded in these grammatical articulations. The question is: what exactly these articulations do? Do they themselves refer to entities of some kind? Does the "-s" on "punches" refer to some mysterious entity – perhaps to kind of function? Is *that* why (A) encodes a proposition

while (B) does not? Are grammatical articulations *referring terms* with the special property that the things to which they refer confer unity on what would otherwise an unordered heap? Or do grammatical articulations not refer to anything? Do they exert their unifying influence in some other way?

§ Let me start out by stating, in extremely rough outline, the position I will defend. Consider the ordered-set <2,3>. The ordinal properties of this set can be coded in, and actually identified with, the properties of some *unordered* set, namely ((2), (2,3)). I will argue that propositions are unordered sets of properties. Thus *Smith punches Jones* is a set of properties. For a proposition to be *true* is specifically for all of its constituents to be *instantiated.* Truth is instantiation. For a proposition to be true is for all the members of a certain kind of set to be instantiated.

Here two worries immediately arise. First, it seems obvious that *Smith punches Jones* and *Jones punches Smith* have the same *constituents,* even though they are obviously different propositions. I will argue that those two propositions *share* constituents – they have in common Jones, Smith, and the relation of punching. But the one proposition has constituents not had by the others; and the ordinal differences between those two propositions are reflected in these differences in their constituents.

The other worry is this: Surely Smith is not a *property;* surely Smith cannot be instantiated. Let me outline what I will say about this. On my analysis, (Ap) is identical with some set S that comprises only things that can be instantiated, and that therefore comprises only *properties*. I will argue that what is a member of S is not Smith *per se* but is rather the property of being identical with Smith. Obviously *that* thing can be instantiated.

No doubt, this maneuver will raise another concern: a concern I will spend a lot of time trying to allay. It will at once be said that what is a constituent of

(Ap) is *Smith* and not the property of being identical with Smith. (Ap) says that *Smith* punched Jones, not that the property of being identical with Smith did so.

I fully agree: what is a constituent of (Ap) is Smith, and Smith – nothing else – is what punched Jones. But we need to be very clear what it means to say that something is a *constituent* of a proposition. As I will argue, when you say that Smith is a "part" or "constituent" of a proposition P, you are not saying that Smith's relation to P is anything like a brick's relation to a house or a wheel's relation to a car. What you are saying is that the truth of P depends in some way (that we will specify) on Smith: the "constituency" relation in question is a relation of *dependence* – it is a purely logical relation, and cannot be modeled on a spatial relation of any kind. Now suppose we identify (Ap) with some set S, one of whose members is the property of being identical with Smith. By doing so, I will argue, we can do justice to the relation that holds between (Ap) and Smith – the relation of dependency that holds between them. So far as it is meaningful, let alone true, to say that Smith is a "part" of a proposition, what is being said is precisely that there exists just the kind of dependency relation we've been discussing. Our construal of propositions as sets of properties, along with our construal of truth as the instantiatedness of the members of such sets, captures that dependency relation to a tee.

There is more to say on this matter. Consider the following two statements:

(i) x is a constituent of a proposition P.
(ii) y is a member of set S.

The sense in which x is a part of P is entirely different from the sense in which y is part of S. x is a part of P only in the sense that the truth of depends on some fact about x. y is a part of S in some other sense. It is true that I identify

(Ap) *Smith punches Jones*

with some set S; it is also true that the property *identical with Smith* is a member of S, and not Smith himself. But the members of S are not the constituents of P. For Smith to be a constituent of (Ap) is specifically for (Ap) to be about Smith -- it is for there to be a dependence relation between Smith and (Ap). Precisely that dependence-relation holds between Smith and (Ap) if we identify (Ap) with S, and then go on the truth of (Ap) with the being-instantiated of all the members of S. So our analysis *is* consistent with the truth of statement that Smith is "constituent" or "part" (Ap), at least on any significant disambiguation of that statement. It is irrelevant that Smith himself is not a member of S. When we are talking about the membership of S we are talking about one thing; when are talking about the constituency of P we are talking about another (we are taking about relations of dependency or aboutness). There is a *connection* between these two sets of facts: facts about the constituency of P – facts about the aboutness relations in question – are *realized by* or *supervene on* facts about the membership of S. But, strictly speaking, the two sets of facts are distinct. We will explore this further in due course.

§ There is yet another worry. Some may object to the idea that there is such a thing as the property of being identical with Smith. It will be said that tallness is a property, and so is wetness – but there is no such thing as the *property* of being identical with Smith.

I think this point of view is wrong for two different reasons. First, statements like "so and so is not identical with Smith" are obviously meaningful – indeed, they are often true -- and it is hard to see how this could

be the case unless there were such a thing as being identical with Smith.

Second, suppose that Smith exists in some world w. Obviously Smith is not a *basic* constituent of w; he is not one of its lower level constituents – he is not a quark or a muon or a single displacement of mass-energy. All of Smith's activities – indeed, his very existence – supervene on lower-level states of affairs. So *if* Smith occupies some part of space-time, that supervenes on much more basic facts – facts about the displacements of mass-energy (or whatever the basic constituents of the world in question are). Nobody is going to deny the existence, let alone the scientific respectability, of the properties involved in those lower level states of affairs – charge, spin, mass, and so on. Nobody is going to say that there is anything suspect or "viciously circular" about believing in these properties. Every fact about Smith, even his very existence, is fixed by facts about those wholly acceptable, lower-level properties. So the property of comprising a Smith is a perfectly respectable property of a world. It is a necessary consequence of there being certain patterns of mass-energy displacement in that world. Given the respectability of these lower-level states of affairs, and given that Smith's existence is a necessary consequence of them, it is hard to see the property of being Smith in any way lacks respectability – it is hard to see why we should deny existence to the property of being identical with Smith.

We might even think of it this way. It is clear that aesthetic and moral properties are realized by lower-level states of affairs: certain patterns of mss-energy displacements realize beautiful works of art, while others don't. Given that Smith's existence is realized by lower-level states of affairs -- by certain patterns of mass-energy – we might even think of Smith as a kind of *property.* If a pattern of mass-energy displacements has the right properties – if, perhaps, it has the right origins, and it has the right internal properties – then it "Smiths".

Of course, there are profound differences between the property of being beautiful and the property of Smithing. A world can comprise many beautiful things; but – leaving aside things like cloning, meiosis, and so on – a world can comprise only one Smith. The reason for this is not hard to identify. Whether a pattern of mass-energy displacements is beautiful doesn't have anything to with its having certain origins – the Mona Lisa would be beautiful in any place or time. By contrast, whether a such a pattern Smiths or not is tied very closely to its conditions of origination. A qualitatively identical pattern that had different origins wouldn't be Smith – it would realize a Smith-duplicate, not Smith himself. But here the important point is that, quite plainly, whatever is the case with Smith – indeed, whether he exists or not – is fixed entirely by scientifically respectable lower-level properties. So it hard to see why we cannot countenance the property of realizing Smith or, therefore, of being identical with Smith.

The analysis given here applies to analytic as well as synthetic propositions, and atomic as well as molecular propositions.

§ I will argue that grammatical articulations *are* what distinguish word-heaps from sentences. I will also argue that grammatical articulations do not typically refer to anything. (In *some* cases they do, but not typically.) If grammatical articulations referred, then they would simply create bigger heaps where, previously, there were only small heaps. Suppose the inflections in (A) refers to entities x, y, and z. In that case, (A) wouldn't be a proposition; it would be the heap: *Smith, Jones, the relation of punching, x, y, and z.* Referring to new things doesn't turn a heap into a proposition; so grammar articulations couldn't perform their unifying function by being referring terms. What grammatical articulations do is to indicate how to generate new properties on the basis of the properties introduced by the referring terms. (I

am treating relations as properties – an n-place relation is a property had by n-tuples.) Consider (A). The referring terms – “Smith”, “punches”, “Jones” – pick out Smith, the relation of punching, and Smith. (For reasons I will give, Smith and Jones are themselves properties.) The purely *grammatical* properties of that sentence – the inflection on the verb, the relative positions of “Smith” and “Jones” – indicate how *new* properties are to be generated from those primitive properties. Those new properties – we will see soon enough what they are – suffice to encode the ordinal differences between *Smith punches Jones* and *Jones punches Smith.*

II. There are, ultimately, three conceptions as to what propositions are.

(i) Functions from worlds (or situations) to truth-values. So *that Plato snores* is the set of worlds where Plato snores or is a function that assigns truth to worlds where Plato snores, and falsity (or no truth-value) to other worlds.

(ii) States of affairs. So the proposition *that Plato snores* is identical with the event of Plato’s snoring (or, perhaps, with a series of such events).

(iii) Arrangements of objects and properties. So *that Smith punches Jones* is some kind of arrangement of Smith, Jones, and the relation of punching, much as <2,6,8> is a structure consisting of three numbers

I believe that (i) and (ii) are false. I believe that (iii) is true as far as it goes, but woefully incomplete. Propositions *are* arrangements of individuals and properties. But we need to know two things. First, what is the relevant kind of arranging? We aren’t talking about spatial arrangement. So what are we talking about? Second, how can an arrangement of objects and properties be *true or*

false? The set (Smith, the relation of punching, Jones) isn't true or false? It doesn't matter how much ordinal structure we introduce: the ordered triple <Smith, the relation of punching, Jones> is no more true or false than its counterpart. We will deal with these problems in due course.

If I am reading him correctly, Wittgenstein gave a clear answer to the two questions we just asked. The sentence "Smith punches Jones" encodes an arrangement of objects and properties: Smith is related to Jones in some way, and the relation in question is of that of puncher to person punched. If reality comprises a state of affairs in which those things are arranged in that way, then the sentence is true. If not, it is false.

Suppose this view is right. In that case, the proposition *that Smith punches Jones* consists of Smith, Jones, and the relation of punching being arranged in a certain way; and they are arranged in the very way in which they would be arranged in the corresponding state of affairs. The sentence is isomorphic with the proposition. The proposition is isomorphic with state of affairs. That is why the sentence represents the state of affairs.

There is an obvious problem with this. If this view is right, then the thing meant by "Smith punches Jones" consists of Smith, Jones, and the relation of punching being arranged in the very way they are arranged in the corresponding state of affairs. How are they arranged in that state of affairs? Smith is punching Jones. So if the view in question is right, the proposition *that Smith punches Jones* consists of Smith standing in the relation of punching with respect to Jones.

But this means that the mere existence of the proposition guarantees its truth. If the proposition *that Smith punches Jones* consist of Smith's punching Jones, then the mere existence of that proposition guarantees that Smith is punching Jones. But the proposition *Smith punches Jones* exists regardless of whether Smith punches Jones. It is obvious that the way in which Smith,

Jones, and punching are arranged in the proposition cannot be identical with the way they are arranged in the corresponding state of affairs.

At the same time, it does seem fair to say that *that Smith punches Jones* consists of Smith, Jones, and punching in some kind of arrangement. But, it would appear, the "picture-theory" seems not to give the right arrangement.

This naturally leads us to position (i): a proposition is a state of affairs. This is not a popular position, since there are two obvious problems with it. First, propositions are *true* or false. But a state of affairs is not true or false, and therefore isn't a proposition (Carnap 1947). Second, (i) makes it impossible for there to be false propositions, unless one takes the heroic of measure of saying that there exist non-existent states of affairs. Some (e.g. Barwise and Perry) have identified propositions with *possible* states of affairs or, at any rate, with sets of such things. This is not so different from position (iii). Before we consider (iii), we should note one other reason why propositions cannot be states of affairs: this other reason is crucial to having a correct conception of what propositions are; it is also crucial to evaluating the Barwise-Perry position.

Consider (Ap). That proposition has a unique decomposition into minimal parts. It neatly decomposes into Smith, the relation of punching, and Jones. (It may have other ingredients, corresponding to tense-markers and the like. Let us ignore these for now.) Notice that, in that proposition, the relation of punching is completely distinct from Jones and from Smith. The proposition *connects* them, of course. But each of these things occurs in its purity: what occurs is not Jones' being tall or angry, but simply *Jones*. The proposition consists of three, neatly separated (though linked) entities.

In light of this, suppose that (Ap) is true. Consider the *state of affairs* that makes it true – imagine Smith punching Jones. Let S be that state of affairs. S doesn't have a structure even remotely like that of (Ap). S's minimal parts are

definitely *not* the relation of punching, Smith, or Jones. First of all, Smith's act of punching is "adjectival" on Smith. It isn't as though Smith is here, and his act of punching is there. Smith's punching Jones is inseparable from Smith and Jones. Further, that act is by no means a *minimal* part of the states of affairs involved. It probably involved a number of subsidiary acts. (Even if it didn't, that doesn't affect our point. Consider the proposition *Beethoven composed the 9th symphony.* The "act" of composing involved innumerable deliberate subsidiary acts.) The same is true *mutatis mutandis* of Smith and Jones. They are not *minimal* parts of anything. They are constituted by innumerable events. If the state of affairs in question has minimal parts, those would be sub-atomic of some kind. But, I think, such events wouldn't really be parts of that event in quite the sense in which Jones and Smith are parts of the proposition *that Smith punches Jones.* It seems more natural to say that the state of affairs, unlike the proposition, doesn't have *any* unique, correct decomposition. In any case, it doesn't have a decomposition even remotely like that of the corresponding proposition.

Also, Smith and Jones are not encountered *in their purity* in the relevant state of affairs. What occurs in the proposition appears to be simply *Jones.* Of course, the proposition goes on to ascribe a property to Jones. But Jones' occurrence in that proposition is unpropertied. What occurs is not Jones' being tall or Jones' being sweaty - just *Jones.* But in any state of affairs, even one that involves Jones' floating in a void, what occurs is not just *Jones*: it is a state of affairs involving Jones. What occurs in reality is Jones' being a in a certain place, having a certain height and weight, animated by certain attitudes or beliefs. There is no Jones *simpliciter.* There are states of affairs involving Jones. Jones is utterly inseparable from them. It is not as though there is *first* a Jones and *then* a state of affairs is constructed out of Jones. For Jones to exist *is* for a state of affairs of some kind to exist. Jones is not accurately thought of

as an isolable constituent of states of affairs. Jones *simpliciter* – the Jones that doesn't occur in a state of affairs – is a pure abstraction. What occurs in reality is not Jones *simpliciter*; what occurs are Jonesy states of affairs. There can no more be "free" occurrences of Jones – occurrences of Jones in isolation of a state of affairs -- then there could be free occurrences red. For these reasons, I think that *metaphysically* Jones is a property of states of affairs. As for what exact kind of property he is, that depends on your other metaphysical views. If you believe the essentiality of origins, then a state of affairs is Jonesy if it has a certain history. If you have a more theologically inclined view, you might say that a state of affairs is Jonesy if it is animated by the right kind of immortal soul. All that is important here is that Jones must be seen as a property of states of affairs, lest we run into insuperable metaphysical problems.

§ It is clear that the *articulations* one finds in a proposition are radically different from those found in the corresponding states of affairs. Ian Hacking refers to this as the "articulation problem". In the proposition *Socrates is tall*, we have one constituent corresponding to Socrates and a totally different one corresponding to tallness. But Socrates cannot be separated from his tallness. Even if you cut off Socrates' legs, you are not separating him from his height: you are only giving him a new height, and he will be as inseparable from his new height as he was from the old one. Propositions separate out what, in reality, are balled together. Propositions and states of affairs have entirely different decompositions. Various thoughts and atoms are constituents of Socrates. But no thoughts or atoms are constituents of the proposition *that Socrates is tall.* So propositions are not states of affairs, and they are not things that are even remotely *like* states of affairs.

However we analyze propositions, we must make sure that the things with which we identify propositions have the same decomposition as the

propositions themselves. It is practically a datum that Socrates is not only a constituent of the proposition *that Socrates is tall*, but a *minimal* and *discrete* part of it. The idea that propositions are states of affairs is inconsistent with that idea. For given any state of affairs in which Socrates is involved, he is absolutely not a minimal part of it, and he is absolutely not neatly separable from it. This fact, I believe, vitiates the idea that a proposition is a set of states of affairs or, what is not so different, a set of worlds. A set of states of affairs has whole states of affairs as ultimate constituents: Socrates himself is no ultimately constituent of such a set. And any given state of affairs cannot possibly have Socrates as an ultimate constituent.

§ This brings us to approach (iii) Consider the set of worlds where Socrates is tall. Socrates isn't a constituent of that set in remotely the sense in which, by all appearances, he is a constituent of *that Socrates is tall*. Now consider a function that assigns truth to each of the worlds just mentioned. The point we just made still applies. On face of it, Socrates isn't a constituent of that function in the sense in which he is a constituent of *that Socrates is tall.*

Of course, attempts have been made to reintroduce constituent-structure into that function – to make it more like the corresponding propositions. Whether these attempts succeed is a question for another time. Even if they do, the possible-worlds approach is still a failure. If a possible world is a set of propositions, then (ii) is viciously circular. If a possible world is a concrete alternative world, then (ii) is committed to all manner of dubious metaphysical and epistemological views. (There is no point in going through them here: they are well documented.) Given this, it would probably be better to abandon (ii).

§ A word of clarification is in order here. I think that it is *true* to say that propositions are functions from worlds to truth-values. But such a statement is

unacceptable *if taken as an analysis*. There are true, and even informative, statements that cannot serve as analyses. Permit me to explain my meaning. First of all, there is surely nothing wrong with talking about hypothetical worlds – such talk is innocent. Let w be a hypothetical world. Some propositions will be true in w and some won't. Let P be a proposition that is true in w. Since P is true in w, we can naturally think of P as associating w with the truth-value *true* (or the property of truth). Let w* be some other hypothetical world, but suppose that P is false in w. P associates falsity with w*. And so on, for any other hypothetical world one might consider. So it is certainly true that propositions associate truth-values with hypothetical worlds, and can thus be regarded as functions from worlds to truth-values.

But propositions *are* functions from worlds to truth-values in the sense of "are" that indicates *predication.* All and only propositions have the property of assigning truth-values to worlds. But if we *analyze* propositions as functions from worlds to truth-values , then our statement is either viciously circular (if the worlds in question are treated as propositions) or is wrong for other reasons (if the worlds are treated as concrete objects).

A comparison might be in order. It may well be *true* that x is the same event as y iff x and y stand in the same causal relations, i.e. if everything that causes x also causes y and if anything that x causes is also caused by y. And such a statement is not only (quite possibly) true, but also informative. But taken as an *analysis* of the concept of an event, that statement is circular: after all, causes and effects ere themselves events.[328]

Here is another example. Nothing that is not a proposition entails anything; and anything that is a proposition entails something. So

(*) all and only propositions stand in entailment relations.[329]

is true. Further (*) is an *informative* statements. But it is not acceptable as an *analysis* of the concept of a proposition. Anything entailed by a proposition is itself a proposition. So (*) amounts to:

(**) all and only propositions entail other propositions.

Taken as an analysis, (**) is viciously circular. I believe the statement *propositions are functions from worlds to truth-values* is informative and accurate. But it is viciously circular if it is taken as an analysis. I believe that PWS is a giant projection of a failure to distinguish the conditions that must be met to provide an analysis of a concept from those that must be met to make a true statement about a concept.

(i) is an out and out failure. (ii) is either viciously circular or is extremely dubious. Further, it is at best an open question whether a set of worlds or a function from worlds to truth-values could have a decomposition at all like that of a proposition. Given this, we should, I believe, try to make some variant of (iii) work.

III. Now it is time to give a positive solution. Let us start with some basic points about propositions. To a first approximation, a proposition is a property of a world. Suppose that *Plato snores* is true in w and false in w*. In that case, w has a property that w* does not. w comprises certain states of affairs that w* does not. The space-time manifold in w has a ripple in it *not* had by the space-time manifold in w*. Kenneth Taylor (1998) speaks of states of affairs as "ripples in the quantum". For Socrates to be snoring is for the quantum to be rippled or indented in a certain way (with the possible qualification that those ripples have to have certain origins).

Being tall or red is a way that individuals can be. Thus tallness and redness

are properties of individuals. A proposition is a way a world can be. It is thus a property of a world. This is why David Lewis (1986) explicitly identifies propositions with properties of worlds. For the proposition *Socrates snores* to be true in w is for w to be a certain way – it is for w's quantum to have certain ripples in it. That proposition is thus a property, and that proposition is true in a world exactly if that proposition is *instantiated* in that world. We can thus identify truth with instantiatedness. A proposition P is a property of a world, and P is true exactly if that property is instantiated.

Here we should deal with what would appear to be a major problem with what we've said so far (to avoid confusion, I'm going to put the objection in the mouth of an imaginary interlocutor):

Let p be the property of being a world where

(*) *Plato punched Socrates*

is true. Let w be a world where the proposition *Plato punched Socrates* is true, and let w* be a world where that same proposition is false. w has p. w* does not. So it is obvious that, for some property x, *Plato punched Socrates* is true in a world exactly if that world has x. This point generalizes without limit: given any proposition P, there is some property P* such that P is true in a world exactly if that world has P*.

But here is the problem. The proposition seems to be more basic than the corresponding property: the property must be understood in terms of the property; and the property's being instantiated must be analyzed in terms of the proposition's being true. This spells the doom of our analysis: for we are trying to analyze the proposition in terms of the property, and the proposition's being true in terms of the property's being instantiated.

For (*) to be true in w, what property must (*) have? It must have p. But what is p? It is the property of being a world such that, in that world, (*) is true. So (*) is more basic than p. p is defined in terms of (*). This means that we cannot analyze (*) in terms of p; and we cannot analyze what it is for (*) to be true in terms of p's being instantiated. On the contrary, we must analyze p in terms of (*), and we must analyze p's being instantiated in terms of (*)'s being true.

This point generalizes without limit: what we just said about (*) is true of any proposition. Consider the proposition

(**) *penguins don't fly.*

Obviously there is some property P such that (**) is true in a world w exactly if w has P. P is the property of being such that (**) is true in it. But we cannot analyze (**) in terms of P, and we cannot analyze (**)'s being true in terms of P's being instantiated. On the contrary, we must analyze P in terms of (**), and P's being instantiated in terms of (**)'s being true. So our account seems doomed to vicious circularity.

I think that this point is entirely wrongheaded. (To simplify discussion, let us leave momentarily aside analytic propositions.) A proposition is never *just* true. It is always true in virtue of something about the world: mass-energy is distributed in the right way; sentient beings are having the right thoughts or feelings. A true proposition registers some fact, and the truth of a proposition supervenes on something more basic. Again consider (*). It is not as though Plato *just* punches Socrates. Plato's punching Socrates supervenes on facts of a much more fundamental kind: Plato punches Socrates in virtue of the fact that

atom a moves in this way at this time, atom b moves that way at that time, and so on. Given two worlds where the same elementary particles move in the same way, either (*) is true in both of them or in neither. So the truth of (*) supervenes on the way in which mass-energy is distributed in the world. That distribution could obviously be described without mentioning Plato or Socrates or the relation of punching. Given any one of the electron-jumps involved in Plato's punching Socrates, that electron-jump could obviously be described without mentioning Plato or Socrates. The same applies to each of the micro-events involved Plato's punching Socrates. Obviously the totality of those micro-events could (in principle) be described without mentioning Plato, Socrates, or the relation of punching. That totality is plainly *more* basic than (*). If (*) is true in a world w, that fact supervenes on the way in which space-time is wrinkled in w. Those wrinkles *underlie* the truth of (*); and they *can* plainly be described without mentioning (*). Thus the property of having those wrinkles is more basic than (*), and the property of having those wrinkles is more basic than the concept of (*)'s being true. So there is no circularity in analyzing (*) in terms of the property a world has if it is wrinkled a certain way, and there is no circularity in analyzing (*)'s being true in terms of a world's having the right wrinkles.

There is more to say in this connection. As we've noted, given any two worlds where mass-energy is distributed in the same way – any two world's whose quanta are rippled in the same way – (*) is true in both of them or in neither: so the truth of (*) supervenes on other facts. But the converse is not true. There can be two worlds where (*) is true but that otherwise differ enormously. There are many different ways that Plato can punch Socrates, and there are many different ways that the surrounding circumstances can vary. So whether (*) is true is fixed by how the quantum is rippled in w. But how the quantum is rippled in w is by no means fixed by whether (*) is true in that

world. This puts it beyond doubt that, if (*) is true in a world w, that is in virtue of w's having some property that is *more basic* than (*): it is in virtue of facts about the basic constituents of w – facts about the way the quantum is rippled in w. The property that a world has if (*) is true in it is thus less basic than the property that a world has if the quantum is rippled the right way in it. So, contrary to what the objector says, there is some property P such that P is more basic than (*), P's being instantiated is more basic than (*)'s being true, and such that (*)'s being true is fixed by P's being instantiated (but not *vice versa*). Thus our account is *not* guilty of vicious circularity, despite what the objector says.

These points generalize. Consider any empirical proposition x. x is not *just* going to be true in w. It is going to be true in virtue of some fact about the basic constituents of w. x will be true in w in virtue of the fact that, as we put it, the quantum in w is rippled a certain way. The property of having that rippling will be *more* basic than x, and a world's having the right rippling will be *more* basic than x's being true. So I think the objector's point is misguided.

Let us sum up. Consider the proposition *Jones punches Smith.* This is true in w iff the "quantum" in w is "rippled" a certain way – if w comprises displacements of mass-energy of a certain kind. So the proposition *that Jones punches Smith* can be thought of as a property of worlds, and that proposition-property is true in a world w iff it is instantiated in w.

IV. But the story just told isn't quite accurate. There are a couple of reasons for this. Let p be the property that a world must have if, in it, Jones punches Smith. p will be instantiated in w exactly if, in w, there is a certain pattern of mass-energy displacements. But exactly these mass-energy displacements will suffice for the truth of infinitely many other propositions. Obviously they will suffice for the truth of every analytic proposition. (Since such propositions are

true under any circumstances, it follows vacuously that whatever happens in a world is sufficient for their truth.) But those same displacements will make true *somebody punches Smith* and *Jones punches somebody* and *somebody punches somebody* and *somebody does something intentionally.*

We could try to get around this. If a mass-energy displacement is enough to make true *Smith punches Jones*, it is enough to make true *somebody punches Jones*, but the converse doesn't hold. So perhaps we could use this asymmetry to distinguish the kinds of properties associated with the patterns of mass-energy displacements.

But this maneuver won't always work. Consider the proposition *x is a circle* and *x is a closed two-dimensional figure of uniform curvature.* Surely these are different propositions; and surely in any possible world, any mass-energy displacement sufficient for the truth of the one is sufficient for that of the other. Of course, some will deny that these are different propositions. But that denial has little independent motivation – it seems an ad hoc way of ameliorating a commitment to a degenerating research program.

Also, the Lewisian picture seems to make analytic propositions disappear. If a proposition is a property of a world, then what a proposition tells you is how a world with that property is different from a world without it. An analytic proposition is one that all worlds have. So, on Lewis' model, analytic propositions don't tell you anything: they might as well not exist.

The problem with the view we've been considering is one we've already encountered. Propositions have decompositional structure. A proposition is not specified merely by its truth-conditions. It is necessary also to give its decomposition The Lewisian analysis just considered doesn't accommodate the internal or decompositional differences between

(i) *Alpha is a figure whose perimeter coincides with the class of all points*

equidistant from a given point in a plane

and

(ii) *Alpha is a closed two-dimensional figure of uniform curvature.*

So it is not enough to say that a proposition is a property of a world. This obviously relates to the fact that analytically equivalent propositions can convey very different information and, what is a consequence, that analytic propositions *do* typically convey information. At the same time, the view just considered contains some profound truths, and our view will be more a development than a rejection of it.

Another illustration may help. Again consider:

(Ap) *Smith punches Jones.*

In this world, if Jones and Smith are to exist, innumerable sub-atomic, atomic, chemical, molecular, cellular…events must occur. In *any* world where Jones and Smith exist, their existence supervenes on the existence of lower-level phenomena. The same is true of the relation of punching: if there is a punching, that supervenes on many other events. So no state of affairs that makes (Ap) true has have Smith or Jones or the relation of punching among its ultimate constituents. Given this, suppose we identify (Ap) with some property P such that (Ap) is true in w exactly if P is instantiated in w. In any world, an instance of P will make true many propositions whose ultimate constituents are *not* Smith or Jones – whose ultimate constituents are tiny displacements of mass-energy. This makes it hard to believe that P itself has a unique decomposition whose ultimate constituents are specifically Smith, Jones, and

the relation of punching. If Jones manages to punch Smith, that is only by the grace of innumerable subvenient states of affairs. So P's being instantiated demands the occurrence of subvenient states of affairs having certain properties. So it is individuative of P, and thus essential to it, that whenever it is instantiated, things *more basic than* Smith, Jones, and the relation of punching are instantiated. For this reason, it cannot be said with confidence that P has a unique, ultimate decomposition into Smith, Jones, and the relation of punching. On the contrary, it seems as though those three things are *not* among its ultimate constituents. This makes P unsuited for identification with (Ap).

All of this assumes that properties themselves even *have* constituents. It is clear that instances of properties have constituents. An instance of the property *chair* has legs and a back. But it isn't really clear if that property itself has constituents. Even if properties do have constituents, it isn't clear if they have them in the sense in which propositions have constituents. Perhaps the property *water* is composed of the property *hydrogen* and the property *oxygen.* But are the last two properties constituents of the first in the very sense in which *Smith* is a constituent of (Ap)? This is, at best, an open question. I will argue that the answer is "no".

Let us sum up. It is pretty clear that Smith and Jones and the relation of punching are constituents of (Ap). It is also clear that (Ap) has a unique decomposition into minimal parts, and that the three things just mentioned are among those minimal parts. But it is *not* clear if P has an even remotely comparable decomposition. It isn't even clear if P has *a* decomposition in the sense in which (Ap) does. Also, P's being instantiated, while sufficient for the truth of (i), is also sufficient for the truth of infinitely many distinct propositions. So it does not seem right to *identify* (i) with P.

§ Here is how I would propose to accommodate all of these facts. (Ap) is a *set* S of properties, *one* of whose members is P. But P is not the only member of S. What are its other members? My answer to this question will initially be alarming.

Let Ps be the property of being *identical* with Smith. Let (Pj) be the property of being identical with Jones. And let (Pp) be the relation of punching. I propose that (Ap) is a set S whose members are: P, Ps, Pp, and Pj. And I propose that for (Ap) to be *true* is simply for all of its members to be instantiated.

One immediate worry about our proposal is this. S doesn't have *Smith* has a member: instead, it has the property of being identical with Smith. By contrast, (Ap) has Smith, not the property of identical with Smith, as a constituent. Therefore, S cannot be identical with (Ap). My response is that it is precisely *in order* to accommodate the fact that Smith is a constituent of (Ap) that we must make the property of being identical with Smith a constituent of S. When we say that Smith is "constituent" of (Ap), that is really a metaphorical way of saying that the truth of (Ap) depends on some fact about Smith: some kind of relation of logical dependence holds between (Ap) and Smith. The sense in which Smith is a "constituent" of a proposition is entirely from the sense in which something is a member of a set. The *membership* of S is not identical with the constituency of (Ap); and in order to do justice to the constituency of (Ap), we *must* assign members to S that are not constituents of (Ap).

Let me flesh out all of these points. Smith is a "constituent" of (Ap). But what does this mean? It doesn't mean that Smith is a part of (Ap) in anything like the sense in which a brick is part of a building. The relation of

constituency, in this context, is purely logical: it isn't mereological. If all the members of S are instantiated, that means

S comprises P. If P is instantiated, that is sufficient for the truth of (Ap). But for reasons earlier discussed, we cannot *identify* (Ap) with P. P doesn't have the right decomposition. One of S's members is a property that, if instantiated, is sufficient for the existence of Smith. Another of S's members is a property that, if instantiated, is sufficient for the existence of Jones. And yet another of S's members is a property that, if instantiated, is sufficient for the existence of an instance of punching.

S has those three things as members and, therefore, as constituents. It is true, of course, that *Smith himself* is not a simple entity. But none of this complexity has any effect on the membership of S. Smith is a member of S – not Smith's leg, not his cells, not the vortex of anxieties that awakens him in the middle of the night. So relative to S, Smith *is* an ultimate constituent – just as he is relative to (Ap). So Smith has the same kind of relation to S that he does to (Ap). Exactly the same point (mutatis mutandis) applies to Jones and to the relation of punching. So, thus far, S and (Ap) appear to have the same decomposition.

But S cannot *just* be identified with the set containing only Smith, Jones, and the relation of punching. All three of those properties could be instantiated *without* Smith's punching Jones – without, indeed, Smith's punching anybody or Jones' being punched by anybody. This is where P comes in. If P is instantiated, that is both necessary and sufficient for the truth of (Ap). So if we include P in S, then (Ap) is true exactly if every member of S is instantiated.

This last point may require elucidation. Any state of affairs sufficient for the truth of (Ap) will be sufficient for the existence of Smith, Jones, and some

punching. So if P is instantiated, so *ipso facto* are Ps, Pj, and Pr. Thus, since the instantiating of P is necessary and sufficient for the truth of (Ap), it follows that the simultaneous instantiating of P1, Ps, Pj, and Pr is also necessary and sufficient for the truth of (Ap).

Here it might be worthwhile to repeat an important point. As we just saw, the instantiating of P *by itself* is necessary and sufficient for the truth of (Ap). But we cannot identify (Ap) with P since P appears not to decompose in the same way as (Ap). If we are claiming that some thing x is identical with (Ap), we must at all costs make sure that x has Smith, Jones, and the relation of punching as ultimate constituents; and if we identify (Ap) with P alone, it is by no means clear if that burden is met. But if we identify S with the set containing exactly P, Ps, Pj, and Pr then, first, (Ap) is true exactly if every member of S is instantiated and, second, the set just described has the same decomposition as (Ap): like (Ap) that set consists of Smith, Jones, the relation of punching, and some complex constituent involving those three.

Let us sum up. (Ap) is true exactly if all the members of S are instantiated. Further, (Ap) and S have the same decomposition. Therefore we might as well identify (Ap) with S and (Ap)'s being true with S's being such that its members are all instantiated.

At this point, there might seem to be a problem for our analysis. We just said that (Ap) and S have *exactly* the same decomposition: and surely they must, if S is to *be* (Ap). But S appears to have a constituent that Ap does not have. S comprises P; but (Ap) comprises only Smith, Jones, and the relation of punching – there is no P. So S and (Ap) have different constituents and are not identical.

I think this is false. I think that (Ap) *does* have a constituent corresponding to P. Smith, Jones, and the relation of punching are not the only constituents of (Ap). That proposition also has a *complex* constituent. Consider the verbal

representation of that proposition: “that Smith punches Jones”. The words “Smith punches Jones” seem to correspond to a single entity, albeit one that is, in some sense, composed of Smith, punching, and Jones. It is this complex entity gives (Ap) its truth-conditions. Smith, Jones, and the relation of punching contribute to (Ap)’s truth-conditions only by way of their involvement in this complex. So while it is true that one of S’s constituents is something other than Smith, Jones, and punching, the same is true of (Ap).

One more application of our analysis might of use. Consider the proposition

(#) *Plato snores.*

Let w be a world where (#) is true, and let w* be one where (#) is false. w is different from w*. The quantum in w is rippled differently: in w, there are events having a certain character and having certain origins; in w*, there are no such events. Let P be the property a world has iff its quantum is rippled in that way. (#) is true in a world exactly if P is instantiated in that world. This might induce one to say that (#) is *identical with* P and that for (#) to be true is identical with P’s being instantiated.

But as it stands, this position is not quite acceptable. (#) has a unique, ultimate decomposition; and Plato and the property of snoring are uncontroversially among those ultimate constituents. (Whether they are *all* of them is another matter: I think not, for reasons I’ll soon give.) Given this, suppose we identified (#) with P. In that case, if our identification is to be correct, then P must have the same decomposition as (#). But it is by no means clear whether this condition is met. If P is instantiated, let us say that there is a P-state of affairs. A P-state of affairs necessarily supervenes on other, lower-level states of affairs: states of affairs involving tissues in somebody’s nasal passages, various interactions among air-molecules, respiratory and

neurological processes, not to mention innumerable sub-atomic processes. Of course, the exact realization of any P-state of affairs may vary from world to world (or even within a world). But what is *not* thus invariant is that if there is a P-state of affairs, that is wholly in virtue of the existence of innumerable lower-level states of affairs meeting certain broad requirements. (It may not matter *exactly* how the air-molecules are distributed. But it matters that they go through *some* kind of passage in a body meeting *certain* requirements, with the result that noise of a *certain* kind is produced: within these limits, there is enormous latitude . But there are limits.) So a necessary condition for P's being instantiated is that various lower-level phenomena are instantiated. The necessary conditions for a property's being instantiated presumably reflect the structure of the property. Given this, it becomes, at the absolute best, an open question whether P decomposes neatly into Plato and the property of snoring. There is the distinct possibility that, if it decomposes at all, it decomposes into things much more basic than those two. Also, as we said before, there is no clear sense in which a property decomposes into others. By contrast, there is one clear sense in which *Plato snores* decomposes into Plato and the property of snoring. For these reasons, it would not be advisable to *identify* (#) with P.

Given this, here is what I would propose. (#) is identical with a set S *one* of whose members is P. The other members of S are Plato and the property of snoring. Remember, Plato is a property. Let Pp be this property, and let Ps be the property of snoring. If P is instantiated, then obviously Ps and Pp are also instantiated – any state of affairs sufficient for Plato's snoring is sufficient for the occurrence of Plato and for an occurrence of snoring. So *P is instantiated* entails *Pp and Ps are instantiated.* Thus, since the instancing of P is necessary and sufficient for the truth of (#), it follows that the instancing of P *and* Ps and Pp are also thus necessary and sufficient.

P is *true* exactly if all of S's members are instantiated. So P is identical with

a set S whose members are Pp, Ps, and P; and P is true in w exactly if all of these properties are instantiated.

S has exactly the right decomposition. Among its ultimate constituents are Plato and the property of snoring. Further, if all the members of S are instantiated, that is necessary and sufficient for the truth of (#). So there is, so far as I can tell, no barrier to identifying S with (#) and to identity (#)'s being true with S's being such that its members are all instantiated.

Now we can say why, on our analysis, it is possible to distinguish between *analytically equivalent propositions.* As we noted, this is a distinction that many theories of the proposition have a hard time accommodating.

There is obviously a difference between the proposition

(#) *that Plato snores*

and the proposition

(##) *that Plato snores and triangles have three sides.*

One of these propositions is about triangles; the other is not. Our analysis easily accounts for the distinction. (#) is a set S of the kind just described. (##) is a set S* that has *some* members in common with S, but that also has other members. Among S*'s members are the property of being a triangle, the property of being a side, the property of being a trio (or of being thrice-instantiated). None of these things is a member of S. So and S and S* have different decompositions, explaining why (#) and (##) have different decompositions.

There is much more to say about the relation between (#) and (##). But it will have to wait until after we've dealt with molecular propositions. But given

only what we just said, it should be clear why our analysis does, whereas the Lewisian analysis does not, account for the decompositional differences between (#) and (##). On Lewis' analysis, (#) is *identical* with some one property P such that P's being instantiated in w is necessary and sufficient for (#)'s being true in w. The trouble, of course, is that P's being instantiated in w is necessary and sufficient not only for (#) but also for (##) and for infinitely many other propositions. The difference between (#) and these other propositions is to be found, at least in part, in their decompositions; and our analysis has a place for those facts.

In light of these points, it should also be clear, at least to a point, why our analysis is able to distinguish between:

(i) *Alpha is a figure whose perimeter coincides with the class of all points equidistant from a given point in a plane*

and

(ii) *Alpha is a closed two-dimensional figure of uniform curvature.*

There is some property P such that if P's being instantiated in w is necessary and sufficient for (i) and (ii) *both* to be true in w. In fact, there is no conceivable ripple of w's quantum that could make the one true without also making the other true. So we cannot *identify* either of those propositions with the property of having a certain kind of quantum-ripple. But since the instancing of P is necessary and sufficient for the truth of each, it is not unreasonable to give it some place in our analysis of those propositions.

Given this, here is what I propose. (i) is identical with a set Si; (ii) is identical with a set Sii. Both sets contain P. But the two sets don't have exactly the same

constituency. Among the members of Si is the property of being a point; but this property is not a member of Sii. Among the members of Si is the property of having uniform curvature; this is not one of the members of Si. Each of (i) and (ii) is true exactly if all its members are simultaneously instantiated. (i) and (ii) are different propositions because, as just indicated, they have different constituencies. The truth of the one is necessary and sufficient for the truth of the other because the instancing of P is necessary and sufficient for the truth of both. The main point is that our set-theoretic analysis distinguishes between analytically equivalent but distinct propositions.

Another example might be in order. The proposition

(*) *a triangle is a triangle*

is not informative. But the following proposition is informative:

(**) *a triangle is a closed figure such that any two of its sides intersect, and such that not all three of them intersect.*

Here it might be appropriate to make a few general remarks. Again consider:

(#) *that Plato snores*

There is reasonable agreement that Plato and the property of snoring are constituents of (#). Our analysis is consistent with that. But our analysis might *seem* inconsistent with another obvious fact about (#). The two constituents just mentioned *combine* to form some larger constituent – one corresponding

to the words “Plato snores” in the verbal expression of (#). Here is the trouble. We identified (#) with a set S whose members are Ps, Pp, and P. On our analysis, P is an *ultimate* constituent of S – it is no less a basic constituent of S than are Ps and Pp. When we look at (#) (or, rather, at its verbal representation), it looks as though the only *basic* constituents of that proposition are Plato and the property of snoring. To be sure, (#) does have a third constituent – one that corresponds to some kind of putting together of the words “Plato” and “snores”. But, by all appearances, that *third* constituent is built out of the other two: it mustn’t be *basic*. But our analysis makes it basic: P is as ultimate and indivisible a constituent of S as either of Ps or Pp. Therefore, one might conclude, our analysis is wrong.

I think that this objection to our analysis involves a radical failure to distinguish propositions from their verbal representations. It also involves a failure to appreciate how much we rely on sheer metaphor when we discuss propositions. Finally, that objection involves a failure to distinguish the psychological facts relating to the grasping of a proposition from the facts relating to the structure of the proposition itself.

There is no doubt that (#) is in some sense “put together out of” Plato and the property of snoring. But there is also no doubt that this putting together has not the slightest resemblance to the putting together of a car or a building. The property of snoring is an abstract object; it cannot be put next to, or on top, or underneath anything. And given any two spatiotemporal entities, no matter how you “put them together” – no matter what kind of glue you use, no matter what kind of forklift you operate – you will not make a proposition out of them. So as for the idea that the third constituent of (#) – the “complex one” – has to result from some kind of juxtaposition of the other two – that idea is a complete non-starter. That third component will not divide into Plato and the property of snoring in anything like the way that table divides into legs and a

plank.

I believe that our analysis actually makes it clear what the *actual* sense is in which Plato and the property of snoring combine to form a third constituent. If P is instantiated, that is *ipso facto* sufficient for the instantiating of Pp (i.e. Plato) and of Ps (i.e. the property of snoring). There is obviously a dependency relation among these various properties. These dependencies are given by entailments. Now properties cannot literally be parts of other properties. A property is an abstract object, and cannot be a part of anything in the sense in which a leg is a part of a body. When we talk about properties being "parts" of other properties, we must surely be talking about relations of *dependence*. These relations are often, though not always, embodied in our verbal definitions: a "triangle" is a "three-sided closed, straight-edged, planar figure". We might say that the property of being a side is a "part of" the property of being a triangle. But we are using the word "part of" metaphorically here, since abstract objects, like properties, don't have parts in any ordinary sense of the word. The *non*-metaphorical meaning here must be this: *if* x has the property of being a triangle, *then* x has the property of having straight-sides. So when we talk about properties being parts of other properties, we must, I think, be talking about relations of dependence or entailment holding among instances of those properties.

That relation does seem to hold among P, Ps, and Pp. If P is instantiated, so *ipso facto* are the other two properties – just as when the property *triangle* is instantiated, so are *side*, *three*, *straight* and so on.

Of course, in the verbal representation of (#), the part corresponding to P is literally composed out of the parts corresponding to Ps and Pp. But that obviously doesn't mean that P itself is literally composed out of those properties. Further, that fact about language supports our analysis, if only weakly. The kind of dependence-relation just discussed is often verbally

expressed through the sort of *verbal* juxtaposition in question: the expression "three-sided closed, planar, straight-edged figure" picks out the property of triangularity. This might make it seem as though the property picked out by that long expression literally *composed* of the properties *straight-edged*, *planar*, and so on: they are literally *parts* of the property of triangularity. But the only sense in which the properties *planar, straight-edged*, and so on, are *parts* of the property of triangularity is that some *dependence* relation holds among these properties: *if* a thing is a triangle, that entails that it is also planar and straight-edged, and so on. So the *verbal* decomposition of "Plato snores" into "Plato" and "snores" corresponds not to a relation of *part*-hood, in any literally sense, but to a relation of dependence. And that is precisely the relation that holds between P and Ps and Pp.

We are inclined to say that Ps and Pp are the only *basic* components of (#), and that any other components are "built out of" those two. The only sense that can be attached to this is: if (#) comprises any third component, then there some kind of asymmetrical relation of dependence—necessitation or entailment -- between that third component and the other two. On *some* interpretation of the expressions "less basic than" and "composed out of", the property picked out by "three-sided closed, planar, straight-edged figure" is less basic than, and is composed out of, the properties picked out by "three-sided", "closed", and so on. But all this means is that *x is triangle* entails *x is straight-edged (closed)* but not vice versa. If we insist on understanding the relation of "composition" in more literal terms, we will inevitably fail. Obviously the aforementioned relation of dependence *does* hold between P, on the one hand, and Ps and Pp on the other. If P is instantiated, that requires that Ps (Pp) be instantiated, but not vice versa. Our analysis of (#) is consistent with that: on our analysis, the component whose verbal expression consists of *both* "Plato" and "snores" stands in exactly that relation of dependence with

respect t the component whose verbal expression is “Plato” and also with respect to the component whose verbal form is “snores”. So our analysis *does* capture the only meaningful sense in which certain components of that proposition are more “complex”, or our “built out of”, others. It is true that, on our analysis, P is an ultimate constituent of (#). But that is in no way inconsistent with standing in the dependence relations just described with respect to Pp and Ps; and, as we’ve seen, to the extent that it has meaning to say that P is a “derived” constituent of (#), and not a basic one, what is meant can only be that P stands in the afore mentioned dependency relations with respect to these other constituents. Since our analysis captures that fact, our analysis captures the only meaningful sense in which certain constituents of that proposition are “built out of”, and are thus “less basic than”, others.

We must take care to distinguish facts relating to the *epistemology* of propositions from facts relating to the structure of propositions themselves. A failure to make this distinction is, I believe, abetted by a tendency to put too much stock in the verbal representations of propositions. When you read the *sentence:*

(#s) “Plato snores”

you “piece together” the meaning of the whole sentence on the basis of the meanings of its parts. This, of course, might leads one to think that the meaning of the whole is put together out of the meanings of the parts in something like the way that a table is put together out of legs and a slap of wood.

When we say that you “piece together” the meaning of the whole on the basis of the meanings of the parts, the only thing this could mean is that you *figure out* the meaning of the whole sentence on the basis of the meanings of

the parts. When we say that you *figure this out*, we can only be saying that you have hit on some dependency relation holding between the meanings of the parts and that of the whole. Once again consider the expression: “three-sided closed, planar, straight-edged figure”. That expression, as a whole, picks out the property of triangularity. On the basis of the meanings of the parts of that expression, plus a small amount of mathematical acumen, you can figure out what the expression as a whole picks out. But this obviously doesn’t that the referent of the whole expression is, in any literally sense, put together out of the meanings of the parts. You can *derive* the meaning of the whole on the basis of the meanings of the parts; and the fact that this can be done embodies the fact that there is a dependency relation between the thing picked out the by whole and the things picked out by the parts. But that is the *only* sense in which the thing picked out by the whole is “composed of” the things picked out by the parts. The *only* sense in which it is meaningful to say that one property is “composed out” of another is to say that there is some kind of dependency relation of the kind described.

To sum up: the fact that we *figure* out the meaning of a whole expression on the basis of the meanings of the parts means only that some (asymmetrical) dependency relation holds between the meaning of the whole and the meanings of the parts. It does *not* mean that the whole is, in any other sense, less basic than the parts. Indeed, in this context, the only meaningful sense that can be attached to phrases like “less basic than” and “is a complex consisting of” concerns relations of dependency. In the only sense of “more basic than” that has any significance in this context, our analysis is consistent with the fact that some constituents of propositions are “more basic than” others. In the only relevant sense of “more basic than”, our analysis captures the fact that Plato (Pp) and the property of snoring (Ps) are more basic than P. Our analysis captures the fact that P’s being instantiated requires Pp’s (Ps’s) being

instantiated, but not vice versa. It is true that, on our analysis, P and Ps and Pp are all equally members of S: from that viewpoint, none is more basic than any of the others; from the viewpoint, there is no stratification. S is a kind of democracy, and P is as lowly as Ps and Pp. But that doesn't matter. The only relevant sense in which P is to be put on a higher level than Ps or Pp is that that instantiating of P instantiated necessitates that of Ps (Pp) but not vice versa. Our analysis captures that fact in spades. Every other kind of stratification or sub-ordination is irrelevant metaphor. So whatever our feelings out it might be, there is nothing logically wrong with identifying (#) with a set in which P, Pp, and Ps are all on an equal footing: their *inequality* – the asymmetry of the entailment relations holding among them -- is inseparable from them, and needn't be buttressed by encasing them in some stratified structure.

These points help us deal with another important fact. Propositions are *structures;* they are *orderings*, not heaps, of constituents. But if we focus too much on language , we are likely to misunderstand the nature of this ordering.

Consider the sentence "Smith punches Jones". The order of the words has nothing to do with the order of the corresponding entities in the proposition. In some languages, the order the nouns is reversed; in others, there are (virtually) no constraints on word-order. The sense in which Smith, Jones, and the relation of punching are "ordered" in

(Ap) *Smith punches Jones*

has nothing to do with order in the spatio-temporal sense.

Once again, the kind of order in question has to do with relations of dependence – relations of entailment. It is true that, in (Ap), Smith has one "position" and Jones has a different one. In

(A*p) *Jones punches Smith*

Smith and Jones have switched positions. What, in non-metaphorical terms, what does this all amount to? We saw earlier that facts about "composition" and "stratification" are really facts about dependence. The same thing is true here facts about the "ordinal" properties of propositions are really facts about entailment relations. (Ap) entails that Smith punches something; (A*p) does not entail this. (Ap) entails that Jones is punched by something; (A*p) does not entail this. And so on. When we say that, in (Ap), Smith is in one "position" and Jones is another, what we are saying is that, if (Ap) is to be true, then Jones must be doing the punching and Smith must be receiving it. We are talking about what is entailed by the truth of the proposition. So we are talking, not about position in any literal sense, but about relations of dependence and entailment. The ordinal properties of propositions are fixed by their inferential properties.

Consider (Ap). We pretheoretically know that Smith is in the "subject" position and Jones is in the "object" position. If we identify (Ap) with something x, then we must be able to retrieve the ordinal information that we pretheoretically have about (Ap) from x. The ordinal properties of (Ap) must be fully "coded" in x. In general, an analysis of what propositions are must fully account for their ordinal properties.

I believe our analysis satisfies this requirement. According to us, (Ap) is identical with a set S whose members are the following properties (remember that individuals are in fact properties)

Ps: Smith
Pj: Jones
Pp: the relation of punching

P: the property that a world has if, in it, Smith is punching Jones.

And for S to be true is for all of those properties to be instantiated. Of course, if all of those properties are instantiated, then P is instantiated. If P is instantiated, there are various consequences: Smith is punching something, Jones is being punched by something, and both punchings are one and the same. We saw a moment ago that the ordinal properties of (Ap) lie in these consequences. So by virtue of identifying (Ap) with S, and by identifying the truth of (Ap) with the instantiating of all of S's members, we have ensured that S codes all the ordinal information characteristic of (Ap).

Here it might be appropriate to point out what are, I believe, some virtues of our analysis. Remember that theory (i), and also Wittgenstein's "picture theory, couldn't account for the existence of false propositions, since they identify propositions with their truth-makers. On our account, the *truth* of a proposition requires the existence of some spatio-temporal state of affairs. But the proposition *itself* requires only the existence of properties: it requires only the existence of *ways* that mass-energy might be displaced.

Let us discuss *molecular* propositions: these are propositions that are formed out of propositions, e.g. *Bob is tall and Sam is short*. I include quantified propositions (e.g. *something is short*) as molecular.

A point about methodology is in order. When dealing with atomic propositions, we defined truth in terms of instantiatedness: a proposition is a set of properties, and that proposition is true iff all its members are instantiated. We must treat molecular propositions in the same way. We cannot define *proposition* and *truth* one way for atomic propositions and another way for molecular propositions. That would be absurd; propositions are propositions; truth is truth. But this need for uniformity requires that, initially, we say a few counterintuitive things.

Consider the proposition *it is not the case that Sam is short* or simply:

(sn) Sam is not short.

The temptation is to say that the constituents of this proposition include *Sam* and *short.* This is correct, but must be qualified. I believe that, like all molecular propositions, (sn) is really a statement *about* a proposition. I believe that Sam is not among the *immediate* constituents of (sn). Rather, some proposition is among those immediate constituents, and Sam is a constituent of *that* constituent. So (sn) is a set. Among the members of that set are the proposition *that Sam is short* and the property associated with the word "not" (we will discuss what this is in a moment). Sam is a member of one of those members, but is not himself a member of (sn).

If the constituents of *that Sam is short* were immediate constituents of (sn), then (sn)'s being true would involve the instantiating of all the constituents of *that Sam is short*. In other words, the truth of (sn) would involve the truth of its own negation.

Consider the proposition:

(ss) *something snores.*

It is tempting to say that the property of snoring is an immediate constituent of (ss). I do not think this is correct. The property of snoring is indeed a constituent of (ss), but not an immediate constituent. It is a constituent of a constituent. (ss) makes a statement about the open proposition *x snores*. As we will see, this open proposition cannot be identified with the property of snoring, even though, of course, that property is a constituent of that open proposition.

Another remark on (ss) is in order. Given what we said before about propositions, it might be tempting to identify (ss) with the class S whose one member is the property of snoring. S would be true iff its one member was instantiated.

Here is the problem with that proposal. We want existence claims, like (ss), *not* to have a radically different structure from other quantified propositions; we want (ss) to be, in at least its broadest outlines, not entirely unlike

(es) *everything snores*

and

(ns) *nothing snores.*

Clearly (ns) and (ss) are cut from the same cloth. Clearly they both express attributions of quantity to some function. Suppose we said that (ns) had the property of snoring as an *immediate* constituent. In that case, given how we analyzed truth, (ns)'s being true would involve the property of snoring being instantiated. So (ns)'s being true would involve its being false.

Also, independently of anything having to do with our analysis, it is clear that (ss) and (ss) – not to mention (es) and other similar generalizations – have the same basic structure. They consist of a quantifier (or, rather, the non-linguistic correlate of a quantifier) and some other thing (which, on investigation, will turn out to be a propositional function). If we *identified* (ss) with a set whose sole member was the property of snoring, then we strip it of this two-tiered structured, and would thus wrongly put it in an entirely different category from its brethren (es) and (ns).

To sum up: molecular propositions are propositions *about* propositions.

The immediate constituents of a molecular proposition are always (i) propositions (or propositional functions) and (ii) properties *of* propositions. The proposition

(ssbt) *Sam is short and Bob is tall*

is *directly* about the propositions *that Sam is tall* and *that Bob is tall.* It is only *indirectly* about Sam: it does make a statement about Sam, but only by way of making a statement about a proposition about Sam (by attributing truth to that proposition).

For these reasons, *it is true that snow is white* and *that snow is white* are different propositions. One is directly about snow. The other is only indirectly about it. These propositions have radically different structures.

Now that we've cleared up these preliminaries, let us give our analysis of molecular propositions. Let us start with:

(jsw) *That John believes that snow is white.*

This is a proposition about a proposition; it attributes a property – being believed by John – to *that snow is white.* Obviously that property and that proposition are constituents of jsw. So jsw is identical with a set S among whose members are that property and that proposition.

Here a refinement is in order. We have identified propositions with sets of properties: it is crucial to our analysis that every component of a proposition be something that can be instantiated. But, as we've analyzed it, the proposition *that snow is white* is not exactly a property; it is a set. (I don't think that properties *are* sets.) In order to provide a uniform analysis, I propose this. What is a constituent of jsw is the property of being identical with *that*

snow is white. That property is instantiated exactly if that proposition *exists.* We must be careful here: the property of being identical with *that snow is white* can be instantiated without that proposition's being *true.* For the instancing of that property, it is necessary and sufficient only that *that snow is white* exist. To sum up, jsw is a set S, and among the members S are the property of being believed by John and the property of being identical with *that snow is white.*

But those cannot be the only members of S. Both of those properties can be instantiated *without* John's believing that snow is white. The property of being identical with *that snow is white* is automatically instantiated. If John believes that grass is green (or even that snow *isn't* white), then the property of John's believing something is instantiated. So the properties in question can be instantiated with jsw being true. S must comprise some other constituent. This other constituent is not hard to identify.

In worlds where John believes that snow is white, the quantum is rippled one way; in other worlds, the quantum is rippled differently. Let P be the property that a world has if, in it, the quantum is rippled in the way just mentioned. So P is instantiated in w exactly if, in w, John believes that snow is white. Among the members of S are P. Of course, if P is instantiated, then so are the properties that are members of S. Obviously if P is instantiated, then John believes something or other (so the property of being believed by John is instantiated). And the property of being identical with the proposition *that snow is white* is instantiated under any circumstances. Thus if P is instantiated, so are the other members of S. And jsw is true exactly if P is instantiated. So jsw is true exactly if all three members of S are instantiated. Thus we identify jsw with S, and we identify S's being true with all its members being instantiated.

The question arises: Why not just say that P is the *sole* member of S, since

P's being instantiated is necessary and sufficient for the truth of S? We've already seen the answer to this. It is pretty clear that jsw comprises a distinct constituent corresponding to the property of being believed by John, and it also comprises a distinct constituent corresponding to the proposition *that snow is white* (and, thus, to the property of being identical with that proposition).If we identify jsw with a set, then we need the membership of that set to clearly reflect these facts about jsw's constituency. If that set's sole member is P, then those facts are eclipsed; the set doesn't have the right decomposition.

In some very broad sense of the word "composed", propositions are composed of their constituents. We saw earlier that this composition-relation is best understood in terms of relations of dependency or entailment holding among those constituents. It is pretty clear that jsw comprises one constituent corresponding to the words "john believes", another constituent corresponding to the words "that snow is white". It is *also* clear that jsw comprises some more "complex" constituent corresponding to the juxtaposition of "John believes" and "that snow is white". If that more complex constituent exists, then so does each of the simpler ones; but both of the simpler ones can exist without the complex one existing. The thing corresponding to "John believes" can exist simultaneously with the thing corresponding to "that snow is white". But that doesn't guarantee the existence of the thing corresponding to the juxtaposition of those expressions. So an asymmetrical relation of dependency holds between that more complex constituent and those less complex ones.

Exactly these points apply to the members of S. If P is instantiated, then so is each of the two properties just mentioned. But both of those other properties can be instantiated with P being instantiated. So there is a dependency relation between P and each of the other two properties just mentioned; and that relation perfectly parallels the one discussed a moment ago – both relations

are asymmetrical in precisely the same way.

Given this, it becomes clear why we must identify jsw with S. S comprises three constituents: one corresponding to the proposition *that snow is white*; another to the property of being believed by John; and a third corresponding to something that, in the relevant sense, combines the two. (Remember that the kind of combining in question consists in the holding of dependency relations: and the right kinds of dependency relation holds among the three things just discussed.) So S has the same constituency as jsw. Further, jsw is true exactly if all three members of S are instantiated. So we can identify jsw's being true with its being the case that all the members of S are instantiated – roughly, we identify truth with instantiatedness. And we identify jsw with S. In this way, we do justice to the decompositional fine-grain of jsw, and also to its truth-conditions.

At this point, I must address what many will find to be a worrisome feature of our analysis. We decided that what is a constituent of

(jsw) *John believes that snow is white.*

is not, strictly speaking, the proposition *that snow is white*, but rather the *property* of being *identical with that proposition.* Obviously this proposal is at least somewhat counter-intuitive. But I think that its counter-intuitiveness is simply another embodiment of some very crude and false theories that we tend to have about the sense in which things are "constituents" of propositions. *That Plato snores* is not built out of Plato the way a building is made out of bricks. (jsw) is not built out of *that snow is white* the way my car is built out of metal parts. As we saw earlier, whenever we relate the notions of "constituency", "composition", and "part" to propositions, we are really talking about relations of logical dependence and entailment. Now on our analysis, the truth of (jsw)

absolutely *depends* on some fact about the proposition *that snow is white*. On our analysis, (jsw) is true exactly if *that snow is white* has the property of being believed by John. In so far as our analysis of (jsw) is consistent with that fact, it fully accommodates the fact that *that snow is white* is "constituent" of (jsw) – for that "constituency" relation just *is* the dependency relation discussed a moment ago. Now *in order* to accommodate that fact, we had to identify jsw with a set S that comprises, not *that snow is white*, but instead the property of being identical with that proposition. But this is in no way inconsistent with the truth expressed by statements like "the proposition that snow is white is a constituent of jsw": for on any interpretation in which such a statement is meaningful, let alone true, that statement amounts to an affirmation of the dependence relation described a moment ago; and, as we just saw, our analysis fully accommodates the existence of that dependence relation. So by identifying jsw with a set S one of whose constituents is the property of being identical with *that now is white*, our analysis actually *captures* the truth of the statement that *that snow is white* is a constituent of jsw, in so far as that statement is meaningful and true.

This point applies to *all* the initially counter-intuitive features of our analysis. We said that *that Plato snores* is a set one of whose constituents is the property of being identical with Plato, and not Plato himself. This might initially rub us the wrong way. We want to say that *that Plato snores* has Plato for a constituent, and not the property of being identical with Plato. I agree: *that Plato snores* has Plato for a constituent. But when we say that Plato is a "constituent" or "part" of *that Plato snores*, what that means – so far as it has a coherent, let alone true, meaning – is that the truth of that proposition *depends* on Plato in some way. That proposition is true exactly if Plato has the property of snoring. *That* is the one and only sense in which Plato is a "constituent" of that proposition. Our analysis is perfectly consistent with that fact. On our

analysis *that Plato snores* is true exactly if Plato has the property of snoring. So our analysis *honors* the fact, so far as it is a fact, that Plato is a "part" of that proposition. Now the way our analysis honors that fact is by identifying *that Plato snores* with a set S one of whose constituents is the *property* of being identical with Plato. So by making that property, as opposed to Plato himself, be a member of that set, we have actually *accommodated* the truth of the statement that Plato is a "part" or "constituent" of *that Plato snores*, at least on any interpretation of that statement in which it is true.

Here it is time to put forth a crucial clarification. We have identified propositions with sets. But the sense in which something is a member of a set is not identical with the sense in which something is a constituent of a proposition. For x to be a constituent of a proposition is for a certain relation of dependence or entailment to hold: the truth of that proposition *depends* on something about x. But that is not the sense in which the number two is member of the set (2,7,9). So, yes, we have identified *that Plato snores* with a set S. And, yes, the property of being identical with Plato is a member of S, but not Plato himself. But that in no way means that our analysis is inconsistent with the fact that Plato is a constituent of *that Plato snores.* The only meaningful sense in which Plato is such a constituent is that the truth of that proposition *depends* on something about Plato. By identifying *that Plato snores* with a set of the kind described, we have *accommodated* the existence of that dependency relation and have thus honored the fact that Plato is a "constituent" of that proposition.

What we said about jsw applies, with only a few obvious changes, to all molecular propositions. Consider:

(sn) Sam is not short.

This attributes falsity to the proposition *that Sam is short.* So among sn's constituents are that property and that proposition. For reasons discussed earlier, in connection with jsw, it would be preferable to say that the property of being identical with *that Sam is short* is a constituent of that proposition. So sn is a set S that has as members (inter alia) the property of being false and the property of being identical with the proposition *that snow is white.* (The property of falsity is instantiated in a world w iff some proposition is false in w. Presumably that property is instantiated in every world. But that needn't concern us here.) Like any other proposition, sn is true iff all the members of S are instantiated. But obviously sn could be false even if both the property of being false and the property of being identical with *that Sam is short* were instantiated. (After all, those properties are instantiated in w if it is false in w that Smith is bald and true that Sam is short.) So S must comprise some third constituent.

In light of our earlier points, this third constituent is not hard to identify. Consider the way the quantum is rippled in a world in which Sam is not short, i.e. in which Sam is of average or above average height. Let P be the property that all and only such worlds have. P's being instantiated in w is necessary and sufficient for sn's being true in w. So P must be a member of S. But we don't want to make P be the only member of S. For now familiar reasons, making P be the sole member of S wouldn't do justice to the decomposition of sn. But we are true to that decomposition if we make P *and* the other two properties (that of being false and that of being identical with *that Sam is short*) be the membership of S. So has the same decomposition as sn. Further, sn is true exactly if all of those three properties are instantiated. So we identify S with sn, and identity sn's being true with S's being such that all its members are instantiated.

Before dealing with quantified propositions, let us deal with conjunctions.

Consider:

(sb) *That Sam is tall and Bob is short.*

First of all, this proposition is true exactly if the two constituent propositions are true. So the constituent whose verbal representation is “and” can be treated as a property that ascribes truth to a pair of propositions exactly if both members of that pair are true. (This analysis is, I will argue later, equivalent to the familiar idea that “and” denotes a function that assigns truth to a pair of propositions exactly if both members of that pair are true.) Let AND be that property. The property AND is instantiated in a world w exactly if some pair of proposition is true in w. (Presumably AND is instantiated in every world. But this needn’t concern us.)

Obviously (sb) has AND as a constituent. For now familiar reasons, (sb) also comprises the property of being identical with *that Sam is tall* and also the property of being identical with *that Bob is short.* Let Pst and Pbs be these two properties. So, for now familiar reasons, sb is identical with a set S whose membership includes AND and Pst and Pbs. But – once again for familiar reasons – those cannot be the only members of S. After all, all of those properties could be instantiated with sb’s being true. So S must contain a fourth constituent. Consider the way that quantum is rippled in a world where it is the case both that Sam is tall and Bob is short. Let P be the property had by all and only such worlds. sb is true in a world w exactly if P is instantiated in w. (Of course, is P is instantiated in w, then so are AND and Pst and Pbs.) If we suppose that S comprises P, then sb is true exactly if every member of S is instantiated. Further, S has the same decomposition as sb – both comprise distinct and separable parts corresponding to the property picked out (in English) by “and”, and also to the proposition *that Sam is tall* and the

proposition *that Bob is short.* So we might as well identify sb with S, and identify the truth of sb with its being the case that every member of S is instantiated.

Similar discussions provide the analysis of propositions like *Bob is short or Sam is tall*, *Bob is short because Sam is tall*, *Bob is short if Sam is tall*, and so on. There are no conceptual obstacles here, though some care must be taken in choosing the constituents corresponding to the connectives.

Let us now deal with quantified propositions, starting with:

(ns) *that nothing snores.*

I think it is important to begin with negative existentials, since they make obvious certain features of quantified propositions hidden by existentials and universals. Following Frege, I see (ns) as saying that every proposition having the form *x snores* is false. For now, I would like permission to treat the form *x snores* (sometimes referred to as a "propositional function" or "open proposition") as an entity in its own right. In a moment, we will discuss what kind of entity it is. (By way of anticipation, it is *not* identical with the property of snoring.) As I see it, (ns) has (inter alia) two constituents: the propositional form just mentioned, and the property of being universally false, i.e. the property had by a thing x if every instance of x is false. So the form *x is a square circle* has that property, since every instance of that form is false. Let us refer to that property as UF. UF is instantiated in a world w iff there is some propositional form such that, in w, every instance of that form is false. (Presumably UF is instantiated in every possible world. But, once again, this needn't concern us.) For now familiar reasons, we must identify ns with a set S whose membership includes (inter alia) UF and the form *x snores.*

There is a (familiar) caveat. We want to identify ns with a set of properties.

On the face of it, the form *x snores* is not a property. So we want to replace it with something closely related to it that is a property. This is easy to do. What is a member of S is not exactly the form *x snores*, but rather the property of being *identical* with that form.

This procedure is exactly analogous to one we used earlier. We decided that (stbs) *that Sam is tall and Bob is short*

is identical with some set S; and decided that what is a member of S is not exactly the *that Sam is tall*, but rather the *property* of being identical with *that Sam is short.* (Let Pss be that property.) We saw that, contrary to first appearances, this procedure actually *accommodates* the fact that *that Sam is tall* is a "constituent" of (stbs). For that procedure *accommodate* the dependence relation that holds between (stbs) and *that Sam is tall*; and, as we saw, that dependence-relation is really identical with the constituency-relation in question. Further our decision to make Pss be a constituent of S enabled us to extend our analysis of truth as instantiatedness to molecular propositions. That decision enables us to capture both the decomposition and the truth-conditions of (stbs): and those are the two important desiderata.

Let us return to
(ns) *nothing snores.*

We want our analysis of (ns) to do justice both to its decomposition and to its truth-conditions. This can be done, if we apply the procedure (mutatis mutandis) just described. We identify ns with a set S that comprises (inter alia) UF and also the property of being identical with the form *x snores.*

Before we continue, a somewhat delicate point is in order. As we will see, propositional forms probably *are* properties. I believe that *x snores* is the property had in common by all and only *that John snores, that Mary snores,*

that Frank snores, and so on. I believe that an *instance* of that form is some *specific* proposition of that form. So an *instance* of *x snores* is *that John snores.*

But *that John snores* is not an *instance* of the property of being identical with *x snores*. An instance of the *latter* property is the form *x snores* itself.

An example may clarify my meaning. An instance of the property *even number* is the number four. But the number four is not an instance of the property *identical with the property of being an even number.* After all, the number four is not *identical* with the property of being an even number. Similarly, *that John snores* is not an instance of the property *identical with the propositional form: x snores*, even though it is identical with an instance of the form *x snores.*

To sum what we've said thus far: I propose. ns is identical with a set whose membership includes the property of being identical with the form *x snores* (let Pxs be that property) and also includes UF. I propose that, by doing this, we can account for the truth-conditions and the decomposition of ns.

But, again for familiar reasons, S cannot comprise *only* UF and Pxs. After all, *those* two properties could be instantiated in a world w where ns was false. Suppose that, in w, the proposition *nothing plays tennis* is true. IN that case, UF is instantiated in w. Further, suppose that Mary snores in w. In that case, Pxs is instantiated in w. (Of course, Pxs can be instantiated in a world even if *nothing* in that world snores. For, it seems to me, even in a world where nothing snores, the property of being *identical* with the form *x snores* exists. But, once again, we needn't dwell on this.) So both UF and Pxs can be instantiated in a world without ns being true. So ns cannot be identical with a set that contains *only* UF and Pxs.

So given that we are identifying ns with a set, that set must contain some third constituent. Again, this is not hard to identify. Consider the way the

quantum is rippled in a world where something snores. Let P* be the property that all and only such worlds have. Now let P be the property that all and only worlds lacking P* have. So a world has P iff its quantum is appropriately *un*rippled. ns is thus identical with a set S that comprises UF and Pxs *and* P. ns is true exactly if P is instantiated. Of course, if P is instantiated, so *a fortiori* are UF and Pxs. It follows that ns is true exactly if P and Pxs and UF are instantiated. But, for now familiar reasons, we don't want to say that S's sole member is P: that wouldn't do justice to the compositional structure of S. So we identify ns with a set S containing all three of the properties just mentioned; and we identify ns's being true with S's being such that all its members are instantiated.

Our analysis made heavy use of the idea that there is such a thing as the propositional *form* x snores? Now we must answer the question: what is that thing? Consider the class of propositions containing all and only *that John snores, that Mary snores*, and so on. The form *x snores* is the property had in common by all and only those propositions. In light of what we said earlier, we can give a precise statement of what this property is. Remember that the proposition *John snores* is identical with a set S whose members are: the property of being identical with John; the property of snoring; and the property P had by world w exactly if, in that world, John snores. An exactly similar analysis applies to *that Mary snores* and *that Ethel snores* and so on. So the property of being identical with the form *x snores* is one and the same with the property of being a set S such that, for some individual x, S contains (i) the property of being identical with x; (ii) the property of snoring; and (iii) the property P had a by world w exactly if, in that world, x snores.

The proposition:

(es) Everything snores

is dealt with in a way exactly analogous to the way we dealt with (ns). First of all, (es) is true if the propositional form *x snores* is "universally true", i.e. is such that each instance of that form is true. Let UT be the property of universal truth. We identify (es) with a set S that contains UT and, for reasons already seen, that also contains Pxs (the property of being identical with the form *x snores*). For now familiar reasons, S must contain a third constituent. Consider the way the quantum is rippled in a world w in which everything snores. Let P be that property. S contains P and UT and Pxs. (es) is true exactly if P is instantiated. If P is instantiated, then so are Pxs and UT. So (es) is true exactly if all of P and Pxs and UT are instantiated. Thus (es) is true exactly if all the members of S are true. Thus we identify (es) with S, and identify the truth of (es) with its being the case that all the members of S are instantiated.

A Post-Kripkean Analysis of Analyticity

1. A priority is a property of propositions. *Triangles have three sides* has that property; *Socrates was wise* does not.

There is a related category -- that of *analyticity*. Analyticity is a property of sentences. It is thought to be a kind of sentential counterpart to a priority. [330] "Triangles have three sides" is analytic. "Socrates was wise" is not.

The obvious way to analyze analyticity is this: a sentence is analytic iff it has an a priori proposition for its meaning.

The correctness of this view was taken for granted prior to Kripke. But Kripke argued compellingly that there are analytic sentences that bear *a posteriori* propositions. [331]

It is, I think, pretty clear that Kripke is right about this. Kripke's own

example of an analytic sentence that expresses an a posteriori truth is controversial. But we will find that there are innumerable *non*-controversial examples of such sentences. So analyticity cannot be defined as the property of bearing an a priori proposition.

Given this, I will try to come up with a correct definition of analyticity. That is the main purpose of this essay. Our analysis is this. A sentence is 'analytic' iff the only *empirical* knowledge needed to know its truth value is knowledge of the semantic rules of the language to which the sentence belongs. A sentence is 'synthetic' iff it is not analytic.

So 'bachelors are male' is analytic, because if one knows what 'bachelor', 'unmarried' etc. mean, one doesn't need additional empirical knowledge to figure out if that sentence is true. And 'bachelors are emotionally unfulfilled' is synthetic because one needs empirical knowledge over and above *semantic* knowledge to know whether that sentence is true or not.

Of course, in many cases, non-semantic knowledge is needed to know that a given analytic sentence is true. Consider the sentence "the sum of the interior angles of a triangle in Riemannian space can be anything up to, but not including, 360°". That sentence is analytic. Now clearly mathematical knowledge is needed to know that it is true. So knowledge of semantic rules will not, by itself, suffice. But the only *empirical* knowledge that is involved in knowing its truth is knowledge of semantic rules: the rest is purely conceptual (mathematical) knowledge.

It is important to note that semantic knowledge is always empirical. I don't need *any* empirical knowledge to know that some mathematical *proposition* holds. But to know that a *sentence* expressing that proposition is true, I do need empirical knowledge: I need to know that "2" refers to *2*, that "+" denotes +, and so on. And such knowledge is empirical.

Some authors have reservations about Kripke's discovery. I will deal

briefly with these reservations (in the Appendix). We will find that, although these authors may well be right, what they say has absolutely no bearing on the meaningfulness or urgency of *our* project: namely, coming up with a definition of 'analyticity' that is *not* 'a sentence is analytic iff it bears an a priori proposition.'

2. Let us now give a Kripke-style argument[332] for the existence of analytic sentences that express a posteriori propositions. Sometimes we use a Fregean sense or a definite description to *fix* the referent of an expression, but not to *give* its meaning. If somebody asks me "who is Gödel?", I can correctly answer by saying: 'he is the guy who proved the incompleteness of arithmetic.' But 'Gödel' does not mean *the guy who proved the incompleteness of arithmetic.* For 'Gödel did not prove the incompleteness of arithmetic', though false, is not self-contradictory. Here I am used a description to *fix the referent* of 'Gödel', not to *give the meaning* of that expression.

The distinction between referent-fixing and meaning-giving can be understood in terms of an ambiguity characteristic of sentences containing the 'means that' operator. Consider the sentence "By 'Gödel' is meant the person who proved the incompleteness of arithmetic." This sentence is 'syntactically ambiguous': the 'means that' ('is meant that') operator can be given 'wide-scope' or 'narrow-scope'. The narrow-scope reading is this:

(*) x is the person who uniquely proved the incompleteness of arithmetic and 'Gödel' means (refers to): x.

The wide-scope reading is:

(**) ‘Gödel’ means: *the person x who uniquely proved the incompleteness of arithmetic*.

(*) fixes the referent of ‘Gödel’ in terms of the description ‘the guy who proved the incompleteness of arithmetic. But (*) does not *give the meaning* of “Gödel” in terms of that, or any other, description. So (*) allows that it is not *self*-contradictory to say ‘Gödel did not in fact prove the incompleteness of arithmetic.’

By contrast, (**) makes it self-contradictory to say ‘Gödel did not prove such and such.’ For (**) *gives the meaning* of “Gödel” in terms of the description “the guy who proved the incompleteness of arithmetic.”

Of course, it is *not* self-contradictory to say “Gödel did not prove the incompleteness of arithmetic.” So (*), and not (**), is consistent with the semantics of ‘Gödel.’

Having made the distinction between meaning-giving and referent-fixing, Kripke made a compelling case that there are analytic sentences that express a posteriori truths. Kripke’s own example has confused many thinkers; so I will use a different one.[333]

We know that there are spies and that there is a tallest among them, but we don’t know who that individual is. Nonetheless we can give him a name: let “Arnold” be the name of that person. Here we are *not* saying let the *meaning* of ‘Arnold’ be *the tallest spy*. We are saying ‘someone x is the tallest spy: let “Arnold” be our name for x.’ So we are using the description “the tallest spy” to *fix the referent* of “Arnold”, not to give its meaning.

Under these circumstances, Kripke says correctly “Arnold is the tallest spy” expresses a *contingent* and a posteriori truth, just as “Gödel proved the incompleteness of arithmetic” expresses a contingent truth. But, Kripke plausibly says, “Arnold is the tallest spy” is *analytic*. The idea seems to be

that to know that ‘Arnold is a spy’ you don’t know have to know anything over and above the semantics of that sentence. (So, in that respect, ‘Arnold is a spy’ is in the same category as ‘there are three feet in a yard’.) Therefore it must be analytic. If Kripke is right, then “analytic” cannot be defined as “having an a priori proposition for its meaning.”

3. It is strange that no one *prior* to Kripke realized that there are analytic sentences that express a posteriori propositions. For there are some obvious examples of such sentences. Consider the sentence “Bush is named ‘Bush.’” The proposition expressed is obviously a posteriori. Bush might have been named “Smith.” There are possible worlds where this is the case. And to *know* that Bush, the person, is named ‘Bush’, one needs to have had quite specific sense-perceptions. I might have lived next to Bush for thirty years, and thus known the man quite well, but never learned that his name was “Bush.” (He might have led me to believe that his name was “Smith.”) To know that Bush is named “Bush”, I have to have had the right kinds of sense-perception (sense-perceptions associated with reliable sources). So the proposition expressed by “Bush is named ‘Bush’” is a posteriori.

But any English speaker knows, wholly on the basis of linguistic competence, that the sentence “Bush is named ‘Bush’” is true.[334] That sentence obviously *must* be true, even though what it says is clearly a posteriori. Anyone who speaks English *ipso facto* knows that “Bush is named ‘Bush’” is true. And this seems to qualify that sentence as analytic.

4. Some interesting examples of analytic a posteriori sentences come up in connection with indexical expressions like ‘I’, ‘here’, and ‘now’.[335] Surely I can know, entirely on the basis of my command of English, that a token of

‘I am here now’, uttered by myself, is true. So it seems not unreasonable to describe such a token as ‘analytic.’ But the fact that I am here now is not something that can be known a priori. *You* don’t know that I am here now, and you couldn’t possibly know it without doing empirical work. Thus, if I were to token “I am here now”, the proposition borne by that sentence would be a posteriori.

What Kaplan says in connection with the “dthat” operator makes it easy to generate analytic sentences that bear a posteriori truths.[336] The sentence “…dthat[the tallest man alive]…” is true iff the person *who in fact* is the unique tallest man alive has the property…x…So if Bob happens to be the tallest man alive, then “dthat[the tallest man alive] smokes” is true iff Bob smokes. Thus “it is a contingent fact that dthat[the tallest man alive] is the tallest man alive” is true. For it *is* contingent that Bob is the tallest man alive.

By contrast, the sentence “it is a contingent fact that the tallest man alive is the tallest man alive” is ambiguous between a true proposition and false proposition.[337]

“Dthat[the tallest man alive] is the tallest man alive” is presumably analytic. No one who knew the semantic rules for the expressions composing it could fail to know that it was true. At the same time, what it asserts is plainly contingent: namely, *Bob* is the tallest man alive. So that sentence is analytic a posteriori.

Given what Kripke says, and also what Kaplan says in connection with demonstratives and the “dthat”-operator, we cannot define an ‘analytic sentence’ as a sentence that expresses a priori proposition. So what *is* the right way to define that term?

5. The relationship is fairly clear. The sentence ‘Arnold is the tallest spy’ is

analytic, but not because it encodes an a priori proposition. For that sentence encodes an a posteriori proposition. And other sentences that encode that same proposition are not analytic.

"Arnold is the tallest spy" is analytic because it is true entirely in virtue of semantic rules. Everything just said about that sentence is true *mutatis mutandis* of "Bush has the name 'Bush'" and "dthat[the tallest spy] is the tallest spy." Such sentences hold true entirely in virtue of semantic rules. (Of course, the semantic rules include not just those of standard English, but those of English *plus* a few extra semantic rules that philosophers have invented.)

Semantic rules can only be known a posteriori. There is no way that one can know *a priori* that the sentence "Bush is named 'Bush'" is true. One needs to know the semantics of English, and such knowledge is empirical. But *given* a knowledge of the semantics of English, one doesn't need further empirical knowledge to know that "'Bush' names Bush" is true. (But one *does* need non-semantic knowledge to know that Bush is named 'Bush.' I could live next to Bush for ten years and not that he was named "Bush".)

This analysis of analyticity applies to paradigm cases of analytic sentences: for example, '2+2=4' and 'Euclidean triangles have interior angles adding up to 180°'. Given a knowledge of the semantics of English, one doesn't need further *empirical* knowledge to know that '2+2=4'. One needs further knowledge, but not further *empirical* knowledge: one needs additional mathematical or conceptual knowledge. So our analysis of analyticity recovers the fact that '2+2=4' is analytic.

To sum up, a sentence S, belonging to a language L, is analytic knowledge of its truth-value requires no empirical knowledge over and above knowledge of the semantic rules of L.

" 'Schnee is weiss' means *snow is white*' is not analytic. That sentence *is* a

sentence of English. (A German sentence is quoted in it; but the sentence as a whole belongs to English.) But to know that it is true, it is *not* enough to know the semantic rules of English; one must *also* knows those of German. Thus our definition does *not* predict that it is analytic. And, indeed, it is not analytic.

"Bush is named 'Bush'" *is* analytic because knowledge of know the semantic rules of English suffices for knowledge of its truth-value. "Triangles have three sides" is analytic for the same reason.

There is a problem we must deal with.

Consider the sentence "Bush is named 'Bush'". That sentence is analytic *provided* that the "object-language" and "meta-language" coincide. If the object language is Chinese, and the Meta-language is English, then that sentence is false, since Bush is probably *not* called "Bush" in Chinese. In fact, the object-language and meta-language must belong to the same *idiolect* of the same language. It would be possible to imagine an idiolect (a dialect) of English in which Bush was not named "Bush" – in which he was named "Bushy" or "Dubya". The sentence "Bush is named 'Bush'" would be false, and thus not analytic, if the object-language belonged to that dialect, while the meta-language belonged to the dialect we are using in this work. But so long as object-language and meta-language *completely* coincide, "Bush is named 'Bush'" is analytic. So it seems to me that, as it stands, your definition of analyticity is deficient; it doesn't take into account the need for meta-language and object-language to coincide.

I disagree. Our definition *does* work precisely as it stands. The objector's point is based on an oversight. (But it does force us to make explicit an

important point.)

The sentence “Bush is named ‘Bush’” is *elliptical* for “Bush is named ‘Bush’ *in English*’ or for “Bush is named ‘Bush’ *in Albanian*” or some other such sentence.

Consider the sentence “the word ‘snow’ refers to snow”. Is that sentence true or false? It is neither; it needs to completed. Depending on how it is completed, it will turn out to be true or false.

(*) “The word ‘snow’ refers to snow *in English*”
(**)“The word ‘snow’ refers to snow *in Swahili*.”

(**) is false. (*) is true.

It is true that there are conceivable dialects of English in which Bush is not named “Bush”. Let D be one of these dialects. Suppose that, in D, the word for Bush is “Dubya”. We’ve agreed that the sentence

(***) “Bush is named ‘Bush’”

is neither true nor false. It must be completed. Consider the following completion:

(****) “Bush is named ‘Bush’ in D”.

That is *not* analytic; for it is false. But why is it false? (****) is a sentence of English – more exactly, of a certain dialect of English (the dialect sometimes referred to as “the Queen’s English”). Of course, (****) *includes* an expression (‘Dubya’) that does not belong to the Queen’s English. But (****) is *itself* a sentence of the Queen’s English. Knowledge of the semantic rules

of that dialect does *not* suffice for knowledge of the truth-value of (****). One *also* needs knowledge of D: one must know that, in D, ‘Dubya’ refers to Bush.

Therefore, our analysis says that (****) is *not* analytic. And, indeed, it is not analytic.

Of course, both D and the Queen’s English are in some loose sense versions of the same language. But they are *different* versions of that language. Really, *strictly speaking*, they are different languages; they have *different* semantic rules. When linguists say that two “languages” are *dialects* of “the same” language, they are really saying that they are very closely related, but still distinct languages. In the present discussion, I am operating on the premise that languages are individuated by their semantic rules: different semantic rules, different languages. This is certainly *a* reasonable delineation of the term “language”, though not the only one.[338] In any case, in the present context, when I talk about “languages”, I am referring to things that are individuated by their semantic rules. So L and L* are the same language iff they have precisely the same semantic rules. By this definition, different dialects *are* different languages.

Analyticity after Kripke

Having laid this groundwork, we can restore the relationship, severed by Kripke, between a priority and analyticity. Whenever a proposition is analytic *in virtue of what it semantically encodes*, it will be analytic *a priori*. “2+2=4” is analytic in virtue of the fact that it encodes an a priori proposition. The same is true of “circles are closed figures” and “any case of knowledge is a case of true belief”. That is why they are analytic a priori.

If a proposition is analytic in virtue of something *other* than what it

semantically encodes, it will be analytic a posteriori: (*) ‘Bush is named “Bush”’ encodes an a posteriori proposition. So, obviously, that sentence is not analytic in virtue of what it semantically encodes. The same is true of “Arnold is the tallest spy” and “dthat[the tallest spy] is the tallest spy”.

Given an analytic sentence, there is an easy way to test whether it is analytic in virtue of what it semantically encodes or not. First, put that sentence in quotation marks. Then put the expression “it is necessary that” in front of the first word of that sentence, but keep it inside the quotation marks. If the resulting sentence is true, then the original sentence was analytic in virtue of what it semantically encoded. If not, not.[339] For example, take the sentence “triangles have three sides.” Replace it with: “It is necessary that triangles have three sides.” This is true; so our test predicts that the original sentence is analytic in virtue of its encoding an a priori proposition. (*Triangles have three sides* is a priori.) And this prediction is correct. (*Triangles have three sides* is a priori. That is why “triangles have three sides” is analytic.)

Now take the sentence: ‘Bush is named “Bush”’. Replace it with: ‘it is necessary that Bush is named “Bush”.’ That sentence is *false.* So our test predicts that the original sentence is *not* analytic in virtue of encoding an a priori proposition. And that prediction is correct (*Bush is named “Bush”* is not a priori). Consider “it is necessary that I am here now”. That is false; so (a token of) “I am here now”, though analytic, is a posteriori. The same is true of: “it is necessary that Arnold is the tallest spy” and “it is necessary that dthat[the tallest person alive] is the tallest person alive.” It is *false* that, necessarily: dthat[the tallest person alive] is the tallest person alive. For Bob might not have been the tallest person alive. It is false that *necessarily* Arnold is the tallest spy. Arnold might not have become spy.

Let us sum up. Given what Kripke said, analyticity cannot be analyzed as

the property of semantically encoding an a priori proposition. The right analysis is approximately this: Given a sentence S belonging to a language L, S is analytic iff the only empirical knowledge needed to establish S's truth is knowledge of the semantic rules of L.

Appendix:

Some authors contest – or at least *seem* to contest – Kripke's contention that there are analytic a posteriori sentences. When we look at what these authors are *actually* saying, it turns out that what they are contesting is something else entirely: nothing they say has any bearing on the question whether a sentence can be both analytic and a posteriori.

The authors I have in mind are Nathan Salmon and Laurence Bonjour. What these authors say is correct as far as it goes. But it doesn't bear against our analysis of analyticity. To evaluate their points, we might as well turn to Kripke's original example, since Salmon and Bonjour frame their discussion terms of that example.

Suppose that at some particular time t, some individual – call him 'Sam' -- decides to let some particular stick serve as a metrical standard. Sam gives that stick a name; he calls it 'S'. And he gives a name to the corresponding unit of measurement; he calls it a 'meter.'

In that case, the sentence

(*) 'the length of S at t is one meter'

is analytic (at least for Sam and those who speak his specific idiolect). *But* there is obviously no way one could know *a priori* the length of some particular stick. So that sentence expresses an *a posteriori* proposition. So (*)

is analytic a posteriori. So I think that Kripke *has* shown that a sentence can be analytic *and* encode an a posteriori proposition.[340]

But this is not quite what Kripke himself says about (*). Kripke makes a different point (at least on the surface): a point that Salmon and Bonjour contest and that is, I am certain, quite false. Kripke asks: 'What is the epistemological status of [*]?' He says, first, that it is *contingent* and, second, that one knows it *a priori*. (In any case, Sam knows it a priori.)[341]

Kripke is right that (*) expresses a contingent truth. For, as Kripke himself says, S *might* have had some other length at t. (If someone had heated it or stretched it, this would have been the case.)

But does Sam know *a priori* that S is one meter long? Salmon says 'no'. Sam knows that S has that length that is (at least in his idiolect) *called* 'one meter.' But, unless Sam has actually *measured* S, or otherwise established S's length in some empirical way, Sam doesn't know what actual length corresponds to that term. He doesn't know whether 'one meter' picks out a length of 42 inches or a length of 38 inches. So what Sam can know without measuring S, or otherwise establishing S's length through empirical means, is merely that S has that length that is *called* 'one meter'. Sam does *not* know what that length is. So unless Sam has actually measured S, or otherwise established S's length by empirical means, Sam does *not* know how long S is, even though, of course, he knows that S's length – *whatever it should turn out to be* – is to be *referred to* as 'one meter.' Bonjour[342] makes essentially the same point, though in a much more condensed way.

In my view, what Bonjour and Salmon are saying is undoubtedly correct. This becomes apparent when we map what they say about Kripke's own example onto one of *our* examples. Consider the sentence

(**) ‘Arnold is a spy.’

(Of course, ‘Arnold’ is to have the semantics that we assigned it earlier in this paper.) First of all, in all the relevant respects, (**) is isomorphic with (*). Unless I know *who* ‘Arnold’ refers to, I don’t know *which* proposition (**) encodes; I don’t whether it is saying that *Fred* is the tallest spy or *Harry* is the tallest spy. But, of course, unless I do empirical work, I can’t know whether it is Fred or Harry or Agatha that is tallest spy. So (**) is *not* a priori; it is a posteriori.

But there is something I *do* know without having to measure the heights of the various spies in the world. I know that, whatever proposition (**) encodes, that proposition is true: I know *that* fact entirely on the basis of a knowledge of the semantics of that sentence. So (**) is analytic. Thus (**) is an analytic sentence that encodes an a posteriori proposition.

Now (*) is, in the relevant respects, just like (**). Unless I know the *actual* length of S, I will have no idea which length is picked out by the expression ‘one meter.’ So unless I do empirical work – unless I measure S (or otherwise empirically establish its length) – I will have no idea *which* length (*) assigns to S – just as I will not know to *which* person (**) assigns the property of being a spy unless I measure the heights of the various spies in existence (or otherwise learn their heights). So, plainly, (*) is an analytic sentence that encodes an *a posteriori* proposition. Contrary to what Kripke says, (*) isn’t a priori at all.

It is possible to render even *more* glaring the error which Kripke is committing. Consider the sentence:

(***) Bush is named “Bush”

No one will deny that this sentence encodes an a posteriori proposition. And virtually no one will deny that it is analytic. If we map what Kripke says about (*) onto (***), then we end up with the assertion that (***) is *a priori* (but contingent). But (***) makes an a posteriori statement. This is what Salmon and Bonjour say; and they are of course right.

> The right response to all this is (as some therapists say) "to take the good and leave the rest." Kripke deserves credit for the new and correct thesis that there are analytic sentences that encode a posteriori propositions. He has given us an example of such a sentence, namely (*).Through his *examples*, if not his pronouncements, Kripke has bequeathed to us an important insight. But Kripke mis-describes his own discovery. (He says that he has discovered a priori contingent propositions. In fact, he has discovered analytic a posteriori sentences.) So in *describing* his own discovery, he said something wrong. But his discovery *per se* is correct. [343]

Extensionality

Abstract: It is generally held that, in some contexts, truth value (and, therefore, meaning) can be changed by replacing referring terms with co-referring terms. Such contexts are described as "intensional" (or "opaque"). We may refer to the view that such contexts exist as "intensionality", and to the view that such contexts do *not* exist as "extensionality". Though it has very solid roots in pre-theoretical intuition, intensionality leads to insuperable semantic, epistemological, and modal problems. In this paper, I argue that,

when we distinguish literal meaning from the pragmatic and epistemological epiphenomena that often surround it, it turns out that a purely extensionalist semantics can easily accommodate all the data that are usually adduced in favor of intensionality. We need only be attentive to some basic facts pertaining to the *epistemology* of meaning – to questions surrounding, not meaning *per se*, but rather to our epistemic access to it. Given the right (extensional) semantics, in conjunction with the right epistemological background, it becomes possible to solve a number of otherwise intractable problems – e.g. Kripke's (1979) Pierre-paradox, and the difficulties reconciling privileged access with the truth of certain kinds of content-externalism.

I. *the received wisdom*

It is generally held that one can change the meaning of a sentence replacing a referring term in that sentence with a co-referring term. "The richest man in America owns a car" has a different meaning from "Bill Gates owns a car". But the latter sentence is what results when a referring term in the former is replaced with a co-referring term.

In this paper, I will try to do two things. First, I will try to show that replacing referring terms with co-referring terms cannot possibly induce a change in sentence meaning. Second, I will try to show what accounts for the strong *appearance* to the contrary.

If cogent, our arguments have far-reaching consequences. We will find that, at least up to a point, the distinction between epistemic and metaphysical necessity collapses: some of the more compelling arguments in favor of this distinction embody a failure to distinguish epistemological facts relating to

our *grasp* of semantics from facts relating to semantics proper. I certainly do not claim that what is said here wholly undermines that important and illuminating distinction. But I think that some of the apparently more compelling grounds for it turn out to embody the confusion just mentioned.

We will also find that what were previously thought to be deep *epistemological* and *metaphysical* points turn out to be projections of what is really a quite innocent *semantic* fact: the fact that, in certain propositional contexts, modal operators can be given varying degrees of scope.

In connection with this, we will rehabilitate the age-old view that one can*not* assent to a proposition and its negation without being irrational. We will find that Kripke's (1979) astute discussion doesn't require an abandonment of this tenet.

We will find that many of the mysteries relating to quantification into "opaque" contexts – contexts governed by (e.g.) modal or epistemic operators – are not mysteries at all, and only seem to be so long as we fail to distinguish, with due care, between facts relating to semantics proper and facts relating to our *grasp* of semantics.

Finally, we will find that, once the distinction between semantics proper and our *grasp* of semantics is made clear, some of the advantages that Russell's theory of descriptions has over its rival (the theory that definite descriptions are referring terms) prove to be quite illusory.

extensionality

In at least *some* contexts, replacing a referring term with a co-referring term preserves truth-value.

(a) "Hesperus is closer to the sun than the Earth"

is true. “Phosphorous” and “Hesperus” refer to the same thing. So

(b) “Phosphorous is closer to the sun than the Earth”.

is true.

Such contexts are described as *extensional.* (a) is an extensional context. So is (b).

If a context is not extensional, it is *intensional.*

Some have held that all contexts are extensional. (Wittgenstein held in this in the *Tractatus.* Carnap held it at certain junctures.[344]) But this has always been a minority view. For there very much seem to be cases where replacing a referring term with a co-referring term leads from truth to falsehood or *vice versa*.

Suppose that (c) is true. It does *not* follow (so it seems) that (d) is true.

(c) John believes that Hesperus is closer to the sun than the Earth.

(d) John believes that Phosphorous is closer to the sun than the Earth

Some striking apparent violations of extensionality involve inter-substitutions of definite descriptions.

(e) John believes that the number that comes after one is less than three.

(f) John believes that the unique n such that n is the cube root of 27 less than the minimum number of sides a solid figure can have is less than three.

(e) may be true. But that doesn't guarantee that (f) is true. But (f) is (supposedly[345]) what results when a referring term in (e) is replaced with a co-referring term.

what I will try to show

There can be no doubt, I submit, that *all* contexts are extensional. Given some recent developments in semantics, this conclusion is inescapable. This is what I will argue.

Obviously there are *apparent* violations of extensionality. But those violations, I will argue, are *merely* apparent.[346]

The present paper will do nothing new. It will merely assemble points made by various semanticists and epistemologists. These points are well-established. And when they are put together, it is clear what must be said: all contexts are extensional.

For some reason, these points haven't yet been assembled. That is why it is the consensus that extensionality is false.

some terminological points

This essay will make heavy use of a certain distinction. It is a commonplace that a sentence (or sentence-token) may *communicate* a proposition other than, or additional to, the proposition that is its literal meaning. The literal meaning of "sometimes life is unfair" is simply: sometimes life is unfair. But that sentence may *communicate* the proposition: *you're not getting the promotion* or *you're fired.* Given the right circumstance, any sentence, no matter what its literal meaning, may communicate any

proposition.

If P is the literal meaning of a sentence (or sentence-token), let us say that S "semantically encodes" P.

What a sentence *communicates* may or may not coincide with what it semantically encodes. So given only that tokens of S communicate P, it doesn't follow that S semantically encodes P – it is an open-question.

Sometimes we confuse the information *through* which we grasp the literal meaning of a sentence with the information (proposition) that is the literal meaning of that sentence. This point is crucial in the present context.

II. *names: our grasp of their semantics versus their semantics per se*

I see a man in the distance; I see that he has such and such properties (e.g. he is leaping about in a decidedly athletic way). I am told his last name is "Smith". I am told

(i) "Smith is a great racquetball player".

Let's discuss the proposition semantically encoded in (i). There is a certain individual O. I saw O. Sometimes O goes by the name "Smith". The proposition semantically encoded in (i) attributes a certain property – that of being a great racquetball player -- to O. That is *all* that proposition does. As long as O has that property, the proposition encoded in (i) is true. So the proposition semantically encoded in (i) is:

(j) *O is a great racquetball player*.

The next day I meet somebody. He has thus and such properties (e.g. he is pudgy and unathletic looking). I shake hands with him. I have a clear and distinct perception of him. I am told his first name is “Bob”. I don’t *realize* that I am shaking hands with the man I earlier saw in the distance. A moment later, I am told

(k) “Bob is a great racquetball player”

Let’s talk about the proposition encoded in (k). Obviously *somebody* must be a great racquetball player if that proposition is to be true. But not just anybody will do. Some *specific* individual must have that property. Who is that individual? It is O of course. And *all* that proposition does is to attribute a certain property (being a great racquetball player) to O. So (k) semantically encodes (j).

But the *cognitive* value of (k) may be very different from the cognitive value of (i). What does this mean? It means that *what (k) tells me* – what I learn about the world – is very different from what (i) tells me.

But how is this possible? After all, (i) and (j) have the same literal meaning. So, presumably, they should both tell me the exact same thing, namely:

(j) *O is a great racquetball player*.

The fallacy here is obvious: what a sentence (or sentence-token) *tells* somebody isn’t always confined to its literal meaning.

To hear a sentence is not to hear a proposition. One must make one’s way from the physical noise one hears to the proposition that it bears.

To hear a sound is not to hear a meaning. One must make one’s way from

the physical noise to the meaning it bears.

When I hear (i), I have to make my way from the noise that I hear to a certain proposition. I have to make my way from the sound "Smith" to a grasp of the person *denoted* by that sound. How do I make that transition?

I know that yesterday I saw *somebody* with such and such properties from a great distance. And I know that, whoever that person is, "Smith" is his name.

The *literal meaning* of "Smith" is simply Smith. "Smith" is a label. Its semantic function is to pick out a certain person. That is all.

But there is a difference between hearing the noise "Smith" and grasping the semantics of that noise. I must make my way from the noise to the semantics. To do this, I must use the knowledge I have of "Smith". What is this knowledge? I know that yesterday I saw *somebody* with such and such properties from a great distance, and that, whoever that person was, his name is "Smith". *This* is the knowledge I must use to transition from simply hearing a noise of the form "…Smith…" to grasping the proposition semantically encoded therein.

So I *grasp* "…Smith…" through an existence claim: *Yesterday, from a great distance, I saw somebody x with such and such properties and:…x…*

Under the circumstances described, I *grasp* the literal meaning of (i) through an existence claim:

(l) *Yesterday, from a great distance, I saw somebody x with such and such properties and: x is a great racquetball player.*

So I *grasp* the literal meaning of (i) though some existence claim. But that literal meaning is no existence-claim.

Now let us talk about how I grasp the literal meaning of (k). The story here is similar to the one just told. Hearing the noise "Bob" is one thing. Grasping

its meaning is another. I must make my way from the noise to the meaning. How do I do that? Well, what do I know about "Bob"? This is what I know. There is a person with thus and such properties (e.g. he is pudgy and unathletic-looking person) with whom I just shook hands; and that person's name is "Bob". Under the circumstances described, I transition from the sound "Bob" to its semantic content via that proposition, that is, by way of that existence claim. So under the circumstances, when I hear a sentence of the form "…Bob…" what is *communicated* to me is:

(m) *A moment ago, I just shook hands with some person x with thus and such properties and:…x…*

So, under those circumstances, when I hear (i), what is *communicated* to me is:

(n) *A moment ago, I just shook hands with some person x with thus and such properties (he was pudgy, had a nervous smile, a weak handshake…) and: x is a great racquetball player.*

It must be kept in mind that the *semantics* of "Bob" is just Bob. "Bob" is a label; not a description; not a quantifier. So the proposition semantically encoded in (k) is simply (j).

But, as we've just seen, I can only *grasp* (j) *through* some existence claim. So, under the circumstances described, (k)'s cognitive value is (n).

(i) and (k) have the same literal meaning. But under the circumstances described I *grasp* (i) and (k) through (l) and (n). (l) and (n) are simply different propositions. Neither so much as entails the other. Thus I can assent to one and *not* assent to the other without being irrational.

Thus it is no mystery why (i) and (k) should, under these circumstances, have entirely different cognitive values. (i) *communicates* (k); and (l) *communicates* (n). Cognitive value is obviously a function of what is *communicated* – what is grasped. And what (i) communicates simply *isn't* what (l) communicates.

(k) and (n) are completely different propositions. Neither entails the other. So I can assent to one without assenting to the other and *not* be irrational.

Thus I can assent to (i) without assenting to (k) (or vice versa) without being irrational. Obviously whether I am being irrational or not is a function of the propositions *communicated* to me. What (k) communicates doesn't entail, and isn't entailed by, what (i) communicates.

Of course, what is semantically encoded in (i) *does* entail what is semantically encoded in (k), and *vice versa.* After all, they both encode the same proposition.

But whether I am rational or not is to be understood in terms of what is communicated to me. Literal meaning is no benchmark of rationality. No one ever maintained otherwise. If somebody gives me conflicting messages through sentences whose literal meaning don't conflict, it would be irrational of me *not* to see a contradiction. And if somebody gives me consistent messages through sentences whose literal meanings contradict each other, it would be irrational of me to see an inconsistency.

Suppose somebody says to Smith "sometimes 2 and 2 make 4, but sometimes they don't – sometimes they make 5". At the level of literal meaning, that sentence is absurd. But if, on those grounds alone, Smith were to reject what I had just been told, he could only be described as less than brilliant.

And this is because whether one is right to reject a sentence is a function not of what it literally means, but of what it communicates. (k) and (i)

communicate different propositions. Neither of those propositions entails the other. So one can assent to the one while rejecting the other without (necessarily[347]) being irrational.

We have thus explained how it is that (i) and (k) can have the same literal meaning while differing dramatically in cognitive value.

summary

A name is just a label. A name's only semantic value is the thing it labels. Co-referring names – names that label the same thing – have the same semantic value. They contribute the very same thing to sentences in which they occur. Replacing "Smith" with "Bob", or "Hesperus" with "Phosphorous", cannot possibly result in a change of literal meaning. But it *can* result in a dramatic change of cognitive value. We've already seen how this happens. Now let us extend these points.

III. *indirect discourse*

We can make statements *about* propositions. We can attribute properties to propositions.[348] For example, we can attribute truth to propositions. ("*That Plato snored* is true." More simply: "it is true *that Plato snored*." Even more simply: "Plato snored".) We can attribute falsity to them. ("It is false *that Plato snored*." "Plato *did not* snore.") There are various other properties we can attribute to propositions.

When we wish to talk about propositions, we typically put a "that" in front of a sentence that has that proposition for its literal meaning. "That Plato

snored” refers to the proposition literally meant by “Plato snored”.

(p) “That Smith is a great racquetball player”

refers to the proposition semantically encoded in:

(i) “Smith is a great racquetball player”.

And

(q) “that Bob is a great racquetball player”

refers to the proposition semantically encoded in.

(k) “Bob is a great racquetball player”

(k) and (i) semantically encode the same proposition. (q) and (p) are terms that refer to the propositions semantically encoded in (k) and (i), respectively. (q) and (p) thus refer to the very same proposition. That proposition is:

(r) *That O is a great racquetball player.*

We’ve already mentioned some of the properties that we attribute to propositions. There are others. Consider the proposition: *that Plato snored.* I can obviously attribute to this proposition the property of being believed by Fred.

Fred believes some propositions. He doesn’t believe others. Nobody will deny this.

If a proposition is believed by Fred, then it has the *property* of being believed by Fred. Nobody will deny *this.*

Consider the proposition *that Plato snored.* Suppose I want to attribute to this proposition the property of being believed by Fred. To do this, I say

(s) “Fred believes that Plato snored”.

It is as clear as day what the literal meaning of (s) is. The proposition semantically encoded in (s) attributes a certain property to a certain proposition. The proposition is: *that Plato snored.* The property is: *Being believed by Fred.*

In light of this, consider the following pair of sentences:

(t) “Fred believes that Bob is a great racquetball player”

(u) “Fred believes that Smith is a great racquetball player”.

(t) attributes a certain property to a certain proposition. The property is: *Being believed by Fred.* The proposition is: *O is a great racquetball player.*

(u) attributes a certain property to a certain proposition. The property is: *Being believed by Fred.* The proposition is: *O is a great racquetball player.*

(t) and (u) semantically encode the very same proposition. But what (t) *communicates* may differ in truth-value from what (u) communicates.

We’ve already made it clear why (k) and (i) communicate different propositions. And we’ve said what those propositions are. Obviously the same principles apply to other expressions. Given what we’ve said about (k) and (i), it is clear what must be said about

(p) “that Smith is a great racquetball player”

and

(q) “that Bob is a great racquetball player”.

Under the circumstances in question, (p) and (q) will have very different *cognitive* values. But they must have the same semantics: they both refer to the proposition *that O is a great racquetball player.*

Thus sentences of the form

“…that Bob is a great racquetball player…”

and

“…that Smith is a great racquetball player…”

will have the same semantics, but very different have different *cognitive* values.

We know that *semantically* “Bob” is a name. But, under the circumstances in question, sentences of the form “…Bob…” communicate propositions of the form:

A moment ago, I just shook hands with some person x and:…x…

So, under the circumstances, “Bob” is *acting* as a certain kind of existential quantifier. In other words, if we focus on what “…Bob…” is

communicating to me, not on what it literally means, “Bob” is acting as a certain kind of quantifier (what linguists call a “generalized quantifier”). Let us refer to quantifiers of that kind as “Russellian quantifiers”.

Consider the sentence:

(t) “Fred believes that Bob is a great racquetball player”

We’ve seen that, although *semantically* it is just a label, “Bob” is *functioning* (pragmatically) as a quantifier.

That quantifier can be given either “wide-scope” or “narrow-scope” with respect to “Fred believes”. So, speaking in terms of what (t) communicates (not what it literally mans), (t) is ambiguous between:

(w) *Fred believes that: a moment ago, I just shook hands with some person x and x is a great Racquetball player*

and

(x) *a moment ago, I just shook hands with some person x and Fred believes that: x is a great Racquetball player*

By exactly similar reasoning,

(u) “Fred believes that Smith is a great racquetball player”.

is (in terms of what it communicates, not what it literally means) ambiguous between:

(y) *Fred believes that: yesterday, from a great distance, I saw somebody x and x is a great racquetball player*

and

(z) *Yesterday, from a great distance, I saw somebody x and Fred believes that: x is a great racquetball player.*

Given that, under the circumstances described, (t) and (u) are ambiguous in the way just described, it is no wonder that they may *communicate* propositions having different truth-values. Obviously (y) could be true even if (w) was false. Obviously (z) could be true even if (x) was false. This much is obvious from basic logic; no recherché results in semantics need be considered.

Let us take stock. If "Smith" and "Bob" really are names, then

(i) "Smith is a great racquetball player".

and

(k) "Bob is a great racquetball player"

must both have the very same proposition for their literal meaning, namely:

(j) *O is a great racquetball player*.

It clearly follows that

(p) “that Smith is a great racquetball player”

and

(q) “that Bob is a great racquetball player”.

are synonymous. They both have the exact same literal meaning; they both refer to the proposition: *that O is a great racquetball player*.

It also follows that sentences of the form

“…that Bob is a great racquetball player…”

will have the exact same literal meaning as sentences of the form

“…that Smith is a great racquetball player…”

In particular,

(t) “Fred believes that Bob is a great racquetball player”

will have the same literal meaning as:

(u) “Fred believes that Smith is a great racquetball player”.

For reasons we’ve seen, what is *communicated* by (t) may yet differ in

truth-value from what is *communicated* by (u).

This is an immediate consequence of the fact that what is communicated by (k) may be very different from what is communicated by (i).

But this difference between (u) and (t) has nothing to do with the literal meanings of those sentences, just as the difference in cognitive value between (k) and (i) has nothing to do with literal meaning.

At the level of semantics, a name is a mere label. Its sole semantic function is to pick out a certain object; it sole semantic value is the thing it picks out. Replacing a name with a co-referring name cannot change the semantics of a sentence. But it can dramatically change cognitive value. The difference in cognitive value is an entirely pragmatic phenomenon: it has to do not with semantics *per se* but with how one *accesses* semantics. Apparent substitution-failures in epistemic contexts are consequences of these pragmatics-generated differences in cognitive value. Those substitution failures are *not* due to a change in semantics. To sum up:

(t) “Fred believes that Bob is a great racquetball player”

must have same literal meaning as:

(u) “Fred believes that Smith is a great racquetball player”.

Given *any* context – modal, moral, causal…-- an argument exactly similar to the one just given shows that replacing a name occurring in that context with a co-referring name cannot possibly lead from truth to falsity, at least not at the level of semantics. This is a simple consequence of the fact that names are mere labels.

IV. *definite descriptions: our grasp of their semantics versus their semantics per se*

But what we have said so far does not establish extensionality. For, at least arguably, names are not the *only* referring terms. Definite descriptions are not names; but they are referring terms. In any case, this is what they *seem* to be. The obvious thing to say about "the first post-master general" is that it refers to the first post-master general (Benjamin Franklin).

In many contexts, replacing definite descriptions with co-referring definite descriptions very much seems *not* to preserve truth-value.

It is easy to establish that these apparent substitution failures are no threat to extensionality.

First of all, it is an open question whether definite descriptions are referring terms. According to Russell they are not. According to Russell,

(i) "…the phi…"

means:

(ii) *exactly one thing x and has phi and…x…*

Russell's theory accounts for the obvious *cognitive* difference between:

(iii) The inventor of bifocals snored

and

(iv) the first post-master general snored.

According to Russell's theory, (iii) means: *exactly one thing x invented bifocals and x snored.* And (iv) means: *exactly one thing x was a first post-master general and x snored.*

So in Russell's view, "the phi" is a quantifier – what we referred to earlier as a "Russellian quantifier".

Thus, as we (in effect) saw a moment ago, Russell's theory accounts for the obvious cognitive difference between:

(v) Fred believes he inventor of bifocals snored

and

(vi) Fred believes the first post-master general snored.

Of course, if Russell is right and definite descriptions are not *referring* terms[349], then the difference in truth-value between (v) and (vi) is no violation of extensionality. Extensionality says: in any context, replacing a *referring* term in that context with a *co*-referring term preserves truth. Extensionality doesn't say anything about what happens when non-referring terms are replaced with non-referring terms; and it doesn't say anything about what happens when referring terms are replaced with terms that *don't* co-refer. So if Russell is right, we can simply disregard definite descriptions in our attempt to prove extensionality.

But, of course, it cannot be taken for granted that Russell is right. It might turn out that definite descriptions *are* referring terms.[350] So to prove extensionality, we must show that *even if* definite descriptions are referring

terms, *apparent* substitution-failures involving them are no threat to extensionality. Let us now do this.

Some (e.g. Strawson) believe that definite descriptions[351] are referring terms. Let us refer to such people as *referentialists.*

Others (e.g. Russell, Quine) thinks that definite descriptions are quantifiers. Let us refer to such people as *Russellians.*[352]

We must begin by noting how much referentialists *agree* with Russellians. The *disagreement* is much less dramatic than the agreement.

Russell says that:

(vii) "the king of France is bald"

means:

(viii) exactly one thing x is a king of France and x is bald.

Russell thus thinks that, if (vii) is to be *true,* then there must be exactly one king of France and that any such person be bald.[353]

For our purposes, the important idea is this. If "the phi" is a quantifier, then its meaning is given by the formula: "..the phi…" means: *there is some x that has phi and…x.…*with the possible qualification that the "there is" may be subject to some kind of context-based restriction.

Referentialists agree *almost* entirely with Russell. A referentialist will agree that if (ix) is to be *true,* then it is necessary and sufficient that there be exactly one king of France x and that x be sick.

Russell said explicitly: "the king of France is bald" means: *there is exactly*

one king of France x and x is bald. Strawson – Russell's arch-enemy – said explicitly: "the king of France is bald" will be *true* if and only if there is exactly one king of France x and x is bald.

So Russellians and referentialists are in perfect agreement as to what must be *true* iff "…the phi…" is to be true.

But Russellians and referentialists *do* disagree about the semantics of "the phi". So *how* do they disagree? Their disagreement can be understood as a disagreement over where to put the "means that" operator in the semantic rule that gives the meaning of the phi.

For Russellians, the rule that assigns meaning to "…the phi…" is this:

(*) "…the phi…"

means

(**) There is exactly one object x that has phi and…x…

For referentialists, the rule that assigns meaning to "…the phi…" is *this*:

If there is exactly one object O that has phi, *then*

(i) "…the phi…"

means

(ii) ...*O*…

(or, equivalent, means: *O has the property/satisfies the function…x…*).

And if that condition isn't satisfied – if *nothing* has phi or if *several* (equally things have phi – then (i) doesn't bear any proposition at all. It is abortive.

So according to the referentialist,

(iii) "the inventor of bifocals snored"

doesn't bear any proposition if nobody invented bifocals or if, on the other hand, several people invented bifocals. But *if* exactly one person O invented bifocals, then (iii) means:

(iv) *O* invented bifocals.

Of course, as it happens, somebody *was* a unique inventor of bifocals: Benjamin Franklin. So O is Benjamin Franklin. So, according to the referentialist, (iii) says:

(v) Benjamin Franklin snored.

According to the view that definite descriptions are quantifiers, (iii) means:

(vi) exactly one object x invented bifocals and x snored.

So there is a significant difference between the Russellian and the referentialists views. After all, (v) doesn't entail (vi), and (vi) doesn't entail (v).

We've already seen that if the Russellian view is the right one – if definite descriptions are quantifiers – then substitution failures involving definite descriptions don't have any bearing on the problem of extensionality.

Let us now see if extensionality is threatened if the referentialist account is the right one.

It most definitely is not. If the referentialist account is right, then

"…the phi…"

means

….*O*…

where O is the object that uniquely has phi. So if the referentialist is right, *all* that "the phi" semantically contributes is a certain *object.*

Of course, the referentialist knows that descriptive information is part of the semantic of "the phi". But the semantic role of that information is to *pick out* a certain individual. That individual is what makes it into the proposition borne by "…the phi…" The descriptive information is not part of that proposition.

Thus, if the referentialist is right, then

(A) "the inventor of bifocals snored"

and

(B) “the first post-master general snored”

semantically encode the very same proposition:

(C) Benjamin Franklin snored.

Of course, the referent of

(D) “that inventor of bifocals snored”

is the proposition semantically encoded in (A). This is not negotiable.

And the referent of

(E) “that the first post-master general snored”

is the proposition semantically encoded in (B). This is not negotiable.

If the referentialist is right, then (D) refers to (C), and so does (E). So (D) and (E) are co-referential.

Thus

“…that the inventor of bifocals snored…”

semantically encodes the proposition

(F)…that Benjamin Franklin snored…

And

(G) “…that the first post-master general snored…”

semantically encodes (F) as well.

Of course,

(H) “Fred believes that the inventor of bifocals snored”

makes a certain statement about the proposition borne by “that the inventor of bifocals snored”. It says that Fred *believed* that proposition to hold. That proposition is: *that Benjamin Franklin snored.* So if the referentialist is right, then at the level of literal meaning, (H) means:

(I) *Fred believes that Benjamin Franklin snored.*

By exactly similar reasoning, it follows that

(J) “Fred believes that the first post-master general snored”

bears (I) as well.

So if the referentialist is right, then (J) and (H) *do* semantically encode the very same proposition. The substitution failure has nothing to do with literal meaning.

Given *any* context – modal, moral, causal…-- exactly similar arguments deal with apparent substitution-failures in that context involving (for example) “the number that comes after one” and “the unique number such

that n+3 is equal to the minimum number of sides a solid figure can have". So *if* definite descriptions really are referring terms, then at the level of semantics, *no* substitution failure is involved in the move from

(K) Little Timmy believes that the whole number that comes right after one is less than ten

to

(L) Little Timmy believes that the unique number n such that n+3 is equal to the minimum number of sides a solid figure can have is less than ten.

If definite descriptions are not referring terms, then they are Russellian quantifiers. And in that case, the difference in truth-value between (K) and (L) is *not* a violation of extensionality. Extensionality doesn't say anything about what happens when quantifiers are replaced with other quantifiers.

There is a glaring difference in cognitive value between (say) (K) and (L) and between (H) and (J). The referentialist must explain why this is so.

He does not, I think, have a hard time dealing with this problem. *Everybody* – no matter what their semantics – agrees that if

(A) "the inventor of bifocals snored"

is to be *true*, then exactly one person x must have uniquely invented bifocals and x must have snored.

When a competent English speaker *hears* (A), he knows that if *nobody* invented bifocals, then (A) will not (literally[354]) be true. And he knows

that if eighty people invented bifocals, then (A) will not be true.[355] So any competent speaker of English knows that, if (A) is to be *true*, then exactly one person must have invented bifocals and that person must have snored.

Thus, inevitably, when a competent speaker of English hears (A), at least part of what is *communicated* to him is: *exactly one person x invented bifocals and snored.*

For exactly similar reasons, when a competent speaker of English hears (B), at least part of what is *communicated* to him is: *exactly one person x was a first post-master general and x snored.*

Again, the referentialist insists that, if (A) is to be *true*, then exactly one thing must have invented bifocals and that thing must have snored. And he insists on the same *mutatis mutandis* in the case of (B). (Up to this point, the referentialist is in perfect agreement with the Russellian.)

Once these facts about (A) and (B) are granted, and once it is granted that competent English speakers have *some* awareness (however minimal and peripheral) of them, it follows inevitably that (A) and (B) will have dramatically different *cognitive* values. So the referentialist, it seems to me, doesn't have the slightest problem accounting for the cognitive difference between (A) and (B).

Basically, the referentialist has no trouble accommodating the fact that (A) will *communicate* (at least in part) some quantified proposition, namely *exactly one person x invented bifocals and snored.* He has no difficulty accommodating the fact the "the phi" *acts* like a Russellian-quantifier, at least in terms of what it *communicates*.

Once it is granted that "the phi" can *act* like a quantifier – if only in terms of what it communicates (as opposed to what it literally means) – then there is no problem at all explaining the cognitive difference between:

(H) “Fred believes that the inventor of bifocals snored”

(J) “Fred believes that the first post-master general snored”.

In effect, we saw this earlier (when discussing the cognitive difference between “Fred believes that Smith is a great racquetball player” and “Fred believes that Bob is a great racquetball player”).

Summary

If definite descriptions are *not* referring terms, then they are quantifiers. If they are quantifiers, then extensionality is not threatened by substitution failures involving definite descriptions. After all, extensionality says nothing about what happens substitutions involving *non*-referring terms are made.

If definite descriptions *are* referring terms, then indeed there are many *apparent* substitution failures. But those failures have nothing to do with literal meaning; they are pragmatic epiphenomena, corresponding to differences in what is *communicated* and not in what is literally meant. Those failures have to do not with semantics per se, but with our *grasp* of semantics.

Finally, someone who is a referentialist about definite descriptions does *not* have difficulty accounting for the existence of those *apparent* substitution failures. This doesn’t mean that the referentialist is necessarily right. But it

does mean that the referentialist isn't necessarily wrong either: given *only* the existence of those apparent substitution failures, it doesn't follow that the referentialist is wrong.

V. *quantifying in*

The following sentence is obviously meaningful:

(*) "John believes that Hesperus is lovely".

And (*) certainly seems to entail:

(**) "There is some x such that John believes x to be lovely".

Quine (1960) says that (**) is meaningless. He says that, in general, sentences which involve "quantifying into" epistemic contexts are meaningless. Why does he think this?

Suppose that (**) *is* meaningful. In that case, the open sentence "John believes x to be lovely" is true for some values of x and false for some values. We can produce a closed sentence of that form by replacing the "x" with a referring term. Such a closed sentence is true just in case the expression that we substitute in for "x" denotes an object having the attribute expressed by "John believes x to be lovely". Obviously "Hesperus" denotes such an object. Since "Hesperus" denotes such an object, so does "Phosphorous". Therefore the following ought to be a true sentence:

(***) "John believes that Phosphorous is lovely".

So *if* (**) is meaningful, then the truth of (*) entails the truth of (***). But, says Quine, the truth (*) does not entail the truth of (***). After all, we are told, (*) and (***) might differ in truth-value. So (**) is not meaningful. And it is not meaningful because it involves a case of quantifying into an epistemic context.

Quine uses an exactly similar argument to show that quantification into modal contexts results in nonsense.

There is a problem with Quine's analysis: (**) obviously *is* meaningful. There can be no doubt that it *is* meaningful to quantify into modal and epistemic contexts. So Quine's argument must involve a fallacy.

What is that fallacy? Quine takes it for granted that the truth of (*) does *not* entail that of (***). We have seen that the truth of (*) *does* entail that of (***). After all, both sentences bear precisely the same proposition. We have also seen why, that fact notwithstanding, (*) and (***) may *communicate* different propositions.[356]

VI. *a note on the Kripke's Pierre-paradox*

Kripke identified an important puzzle through the following clever thought-experiment. Pierre grows up in Paris. He speaks only French, not a word of any other language. He's seen French movies, and read French books, about London. On the basis of this information, plus his knowledge of the French language, he very rationally comes to assent to the French sentence:

(i) "Londres est jolie".

Of course, Pierre understands (i); he knows what it means. And given the

information at his disposal, he is quite right to assent to it.

The English translation of that sentence is

(ii) “London is pretty”.

Of course, both sentences encode the proposition

(iii) London is pretty.

One day, Pierre is kidnapped. His kidnappers keep him hostage in a rundown, ugly part of London. Pierre learns English. But he learns it *not* by having English sentences translated it into his native French. Rather, he learns it as a person learns a first language.

On the basis of the information available to him, plus his newly acquired knowledge of English, Pierre comes to assent to the English sentence:
(iv) “it is not the case London is pretty”.

Pierre understands (ii). Given the information at his disposal, he is quite right to assent to it. It is the rational thing to do.

Here is the problem. Pierre rationally assents to (i). He also rationally assents to (iv). But the proposition semantically encoded in (iv) is the negation of the proposition encoded in (i). So Pierre assents *both* to a proposition *and* its negation *and* he is not in the slightest bit irrational.

But how is this possible? For to be irrational just *is* countenance some inconsistent set of propositions. In any case, this is the age old presumption.

Further, because of the way Kripke set up the example, we cannot, so it seems, say that use of the sense-reference distinction will solve the problem. Being translations of each other, “London” and “Londres” cannot differ in

sense. For the same reason, we cannot say that “London” and “Londres” are Russellian-quantifiers. For *even if* Russell is right to say that “London” and “Londres” are quantifiers, this won’t help: for being translations of each other, they must denote the *same* quantifier.

So Kripke has shown that *even if* Frege’s views and Russell’s views concerning the semantics of proper names are correct, this does not settle the problems those views are meant to settle. For the very paradoxes they are meant to solve recrudesce when we *know* that the referring terms in question have quite the same semantic content. After all, no matter what you think about referring terms, you must agree that “London” just *is* the translation of “Londres”. Any sense or quantificational structure that the one expression has will be had by the other. So Kripke has made a point that apparently (and, I think, actually) devastates some apparently very reasonable semantic views.

More importantly in the present context, Kripke has also identified an important puzzle: Pierre *rationally* assents to sentences that he correctly understands and that negate each other. I would now like to show that some of our earlier remarks might give us some leverage on this puzzle.

The semantic value of “London” is London; the semantic value of “Londres” is London. But the *meaning* of “London” must be *given* to Pierre through some proposition that is at least somewhat descriptive. Under the circumstances described, that proposition is something like: there is this awful place x, where I am currently imprisoned, where there is rotting garbage in the street…and “London” is the English label for x.

So even though “London” means only *London*, and has no descriptive content, nonetheless what (for example) “London is pretty” communicates to *Pierre* is something like:

(v) there is this awful place x, where I am currently imprisoned, where there

is rotting garbage in the street, and the people are surly… and x is pretty.

By exactly similar reasoning, what "Londres est jolie" communicates to Pierre (not what it semantically encodes) is something like:

(vi) there is this wonderful place x I've read about on thus and such occasions, where there are Beefeater guards, and a big clock-tower…and x is pretty.

To close the argument, we have to deal (once again) with a subtlety relating to scope. Suppose Pierre hears somebody say "London is pretty." Pierre violently recoils: "No! London is *not* pretty. Not one bit!" We've agreed that, *to* Pierre, under the circumstances described, "London is pretty" communicates (v). But this leaves it open what "London is *not* pretty" means *to him.* For we can give the negation either wide or narrow-scope. The wide scope is:

(vii) it is not the case that: *there is this awful place x, where I am currently imprisoned, where there is rotting garbage in the street, and the people are surly… and x is pretty.*

The narrow scope reading is:

(viii) there is this awful place x, where I am currently imprisoned, where there is rotting garbage in the street, and the people are surly… and it is not the case that: x is pretty.

It seems to me that, under the circumstances, (viii) is more likely than (vii) to be the proposition that Pierre associates with “London is not pretty”.

But here is the main point. The conjunction of (vi) and (vii) is not a self-contradiction. Neither is the conjunction of (vi) and (viii). The proposition Pierre associates with “Londres est Jolie” is (vi). And the proposition he associates with “London is not pretty is either (vii) or (viii). Given the propositions that, under the circumstances, Pierre associates with those two sentences, he is obviously in no way guilty of self-contradiction. Not even the *appearance* of self-contradiction arises. We don’t need to posit “propositional guises” or other dubious entities.[357] We explain how it is that Pierre can be rational – even though he is assenting, with understanding, to a given sentence *and* to a sentence that is its negation -- while holding onto the extremely reasonable view that “Londres est jolie” and “London is pretty” semantically encode the same proposition.

Importantly, our analysis does *not* require us to forfeit the age-old view that irrationality consists in assenting to inconsistent propositions. For (viii) is *not* inconsistent with (vi), and neither is (vii). Kripke’s story seems to make it *de rigueur* to revise, or qualify, the view that to be irrational is to assent to conflicting propositions. But, it would seem, this is not necessary. We need only be sensitive to the distinction between what sentences literally mean (semantics), on the one hand, and how we go about *interpreting* sentences, on the other.

VII. *a qualification to the distinction between epistemic and metaphysical possibility*

In these post-Kripke (1972) days, things like the following are often said:

(1) given only that *Hesperus is lovely* is true, it is nonetheless *epistemically* possible that *Phosphorous is not lovely* is true.

A related (if not identical) claim is that

(2) The proposition *Hesperus is lovely* doesn't entail *Phosphorous is lovely*.

At the same time, it is held that, if Hesperus is lovely, then as a matter of *metaphysical* necessity, Phosphorus must also be lovely.

So, it is concluded, there is a distinction between epistemic and metaphysical possibility/necessity.

Far be it from me to attempt to undermine this excellent distinction. But I think that, in light of some remarks earlier made, we have reason to scrutinize *some* (though certainly not all) of the reasons adduced in favor of this distinction.

Consider the *sentences* (as opposed to the propositions):

(3) "Hesperus is lovely."

(4) "Phosphorous is lovely."

"Hesperus" and "Phosphorous" both label the same thing. Let O be that object. And those two expressions are nothing *but* labels. So, in terms of their semantics, (3) and (4) should be synonymous. Both must encode the proposition:

(5) O is lovely.

Thus (1) amounts to the following claim:

(6) given only that *O is lovely* is true, it is *epistemically* possible that *O is not lovely* is true.

But (6) is just plain false.

At the same time, it must be granted that what is *communicated* by (1) is, or can be, true. How is this possible. We've seen that sentences of the form "…Hesperus…" will typically *communicate* (not semantically encode) some existence claim, the same being true of "…Phosphorous…" So "Hesperus is not lovely" might communicate :

(7) there is some object x that appears at such and such time and such and such part of the sky and has such and such other properties…and x is lovely.

And "Phosphorous is lovely" might communicate something like:

(8) there is some object y that appears at thus and such time and thus and such part of the sky and has thus and such other properties…and y is not lovely.

Obviously it is epistemically possible that (7) and (8) should both be true.

But we still have a problem. For the following seems to be true.

(9) Given that Hesperus is lovely, it is *metaphysically* necessary that Phosphorus be lovely.

How do we explain this on our current account? It is just a matter of where to situation the relevant operators (the "it is epistemically necessary that" and the "it is metaphysically necessary that") in the relevant existence-claim.

First let us conjoin (7) and (8):

(10) There is some object x that appears at such and such time and such and such part of the sky and has such and such other properties…and x is lovely; and, further, there is some object y that appears at thus and such time and thus and such part of the sky and has thus and such other properties…and y is not lovely.

When we say that (1) is true, what we mean is something along the lines of:

(11) **it is possible that**: There is some object x that appears at such and such time and such and such part of the sky and has such and such other properties…and x is lovely; and, further, that, there is some object y that appears at thus and such time and thus and such part of the sky and has thus and such other properties…and y is not lovely.

So (1) means we can put an "it is epistemically possible that" operator in front of some conjunction of existence claims.

When we say things like (9), what we mean is something along the lines of:

(12) There is some object x that appears at such and such time and such and such part of the sky and has such and such other properties; and, further,

that, there is some object y that appears at thus and such time and thus and such part of the sky and has thus and such other properties, and **it is necessary that**: if x is lovely, y is also lovely.

Obviously (12) is perfectly correct, given that (so to speak) x *is* y. At the same time, (11) is perfectly correct, since the one existence claim obviously has no analytic relation to the other. When we say things like:

(13) It is epistemically, but not metaphysically, possible that Hesperus should be distinct from Phosphorous

we must mean something like the conjunction of (11) and (12). This analysis accommodates the following facts:

(fact #1) “Hesperus is lovely” and “Phosphorus is lovely” semantically encode the same proposition.

(fact #2) Given fact #1, (1) semantically encodes the following totally incoherent statement: given only that *O is lovely* is true, it is *epistemically* possible that *O is not lovely* is true.

(fact #3) What is *meant* by (1) – not necessarily what is semantically encoded in it -- is surely, on some reading, quite right

The analysis we’ve given – involving one’s grasping singular positions through existential information, and the subsequent latitude one thereby has as to where to situation modal operators in conjunctions formed from those

existential positions – gives us a handy way of dealing with facts 1-3. The other side of the coin is: if we reject that analysis, that it becomes not entirely easy to deal with those same facts.

I would like to end with a conjecture. These days, it is, of course, a common place that epistemic necessity and metaphysical necessity do not coincide. I would propose that, in at least some cases, the right analysis of these notions is as follows.

Consider some sentence that, supposedly, expresses a proposition that is metaphysically but not epistemically necessary, say:

(*) "Hesperus is Phosphorous.

We've argued that this sentence is cognitively non-trivial because of the *existence* claim through which one grasps its content. When we say that (*) is *metaphysically* necessary, what we are saying is that the result of giving *narrow-scope* to the necessity operator results in a truth. When we say that (*) is not *epistemically* necessary, what we are saying is that the result of giving *wide*-scope to the necessity-operator results in falsity.

The existence-claim through which accesses the proposition encoded in (*) is something like this:

(**) There is a celestial body x that uniquely appears before any others in the evening sky and there is a celestial body y that uniquely disappears after any others in the morning sky, and x=y.

Give the necessity-operator narrow-scope (or, more exactly, narrow*est* possible scope). The result:

(***) There is a celestial body x that uniquely appears before any others in the evening sky and there is a celestial body y that uniquely disappears after any others in the morning sky, and **necessarily: x=y.**

This is a true claim, and it captures the sense in which (*) is necessary while being non-trivial.

Now give the necessity-operator wide-scope (or, more exactly, widest possible scope). The result:

(****) **Necessarily**: There is a celestial body x that uniquely appears before any others in the evening sky and there is a celestial body y that uniquely disappears after any others in the morning sky, and x=y**.**

This is a false claim, for the reasons Kripke gives. Were astrodymamics a bit different, there might not have been a single thing that was *both* the first celestial body to appear in the evening sky *and* the last celestial body to disappear from the morning sky.

A sentence S expresses a proposition that is *epistemically* necessary iff we can give widest possible scope to the necessary-operator in the existential information through which we access the semantic content of S.

(*****) “Hesperus is Hesperus”.

As with any sentence encoding an object-involving proposition, we grasp it through existential content. This content is something like:

(******) There is a celestial body x that uniquely appears before any others

in the evening sky, and there is a celestial body y that uniquely appears before any others in the evening sky, and x=y.

Obviously if we give the “necessarily” operator widest possible scope, we get a truth:

(*******) **Necessarily**: There is a celestial body x that uniquely appears before any others in the evening sky, and there is a celestial body y that uniquely appears before any others in the evening sky, and x=y.

In general, the difference between epistemic and metaphysical necessity seems to reduce, in at least some cases, to where the necessity operator can be put in the *existential* propositions through which we grasp singular proposition. P is metaphysically necessary means: The necessity operator can be given narrowest possible scope. Finally *S expresses an epistemically necessary proposition* means: We can give the “necessarily” widest possible scope in the proposition through which we access the semantic content of S. [358]

VIII. *some needed refinements to our analysis: epistemological foundations*

As it stands, our analysis is vulnerable to an important objection. (Henceforth objections to my views are placed in the voice of an imaginary interlocutor.)

Here is the story you tell.

On a certain occasion t, I meet a pudgy, unathletic looking person. So my

perceptual access to that individual is mediated by existential information like: *on occasion t, there was a certain pudgy, unathletic person, with a weak handshake and a nervous smile, and such and such other properties.*…I am told that person's name is "Bob". Consequently, even though the semantic content of "Bob" is just some individual, my *access* to that content is mediated by the kind of existential content just described. The result is that "Bob is a great racquetball" *communicates* a certain existence claim. The same story applies (mutatis mutandis) to all other names.

This is the foundation of your account of why the cognitive value of "Bob is a great racquetball player" may differ in cognitive value from a synonymous sentence like "Smith is a great racquetball player".

But there is an obvious problem with this account. There are people, and things, that I've known for years. I have not the slightest recollection of the circumstances under which I initially met them. Consequently, I have not the slightest recollection of the existential information involved in my first encounters with them. That information may have been involved in my thoughts about them for a little while after meeting them – perhaps a day or so, perhaps less. But that information quickly drops out. And yet I have absolutely no trouble having thoughts about those people. And I have no trouble accessing the semantic content of terms denoting them. Obviously I *do not* access that semantic content via the existential information you talk about; for, as we just saw, that existential information all falls away quite quickly after the initial encounter. All that is left is a pure object-involving concept: I grasp Bob, but *not* via all the existential nonsense you speak of; I just grasp *him*. And all the descriptive information vanishes, like the morning fog: leaving your analysis without a leg to stand on.

The objector is right on one thing. The "start-up" existential information certainly does drop out. I may not remember, even sub-consciously, the circumstances under which I met (say) Sam. But this absolutely does *not* mean that, at any point, I have some kind of "pure" conception of the person, i.e. one that is not paved with existential information. Indeed, if such "pure" conceptions were possible, it is hard to see how it could be possible for Cicero/Tully (Hesperus/Phosphorous) cases to arise. But, I fully grant, the objector's point definitely requires that we supplement our analysis.

Suppose my first encounter with Sam is through existential information like: *there is guy wearing a green shirt and a bowler hat, whom I am now (on occasion t) meeting and.*…and, consequently, that such information is, at least initially, implicated in my interpretetings of "Sam".

I meet Sam again, four days later. And I do so with knowledge: I *recognize* him. What is happening here? My recognizing him obviously doesn't consist in my becoming aware of some bare identity; it doesn't consist in its being the case, that for some object O, I suddenly become aware that *O is identical with O.* No doubt we do recognize the truth of such bare identities. But that is not what *recognition* consists in. Recognition is not so informationally impoverished.

The story I would tell is this. I realize that there was some chap whom I met a few days earlier, on occasion t, who had such and such features (e.g. he was bearing a bowler hat)…and I also realize that, right now, on occasion t*, there is some chap, who has thus and such properties (e.g. he is wearing a baseball cap)…and I also realize that there is some *one* individual who uniquely satisfies both of these existential propositions.

Let E be the first existence claim, and let E* be the second.

It thus seems that recognizing something consists in seeing that some one thing uniquely satisfies each of *two* existence claims. This means that E -- the

"start-up" existential information -- can be dropped. My *recognizing* Sam involves my seeing that some one entity satisfies both E and E*. Having made this recognition, I am at liberty to drop E. My awareness of Sam will be sustained, for the time being, by my knowledge of E*. And, in its turn, E* may be dropped and replaced by a third existence claim.

So what sustains a conception of an entity is knowledge of a series of interlocking existence claims. Any given one of these claims may be dropped, and probably will be. But at any juncture in the career of one's having a conception of a thing, that conception involves knowledge of *an* existence claim. And this point, as we've seen, enables us to account for the cognitive difference between "Hesperus is lovely" and "Phosphorous is lovely".

Here I should deal with an objection to this theory:

You have *ex nihilo* produced this strange theory of conception as being sustained by knowledge of interlocking existence claims. And you claim that recognition consists in seeing that some one object satisfies two existence claims. But all of this seems very *ad hoc* – something you've confabulated to get your bankrupt theory off the hook.

I think that this comment has force only to the extent that we've been jaded by dubious theories of sense-perception and, more generally, of objectual awareness.

When you see Bob, embedded in your sense-perception is existential information. You cannot get rid of that existential information without eliminating the perception. This suggests that that existential information is constitutive of the content of the perception. Granted, Bob himself is (part of) the representational content of that perception. But he is given to you via this

existential information.

So when you see Bob on some occasion t, the content of that perception is given by some bit of existential information like: *there is some chap whom I'm now seeing, wearing a bowler hat…*

You see Bob again, on occasion t*. Again, the content of your perception is given by some existence claim: *there is some chap I'm seeing, wearing a baseball cap…*

On both occasions, the content of your perception is existential.

Suppose that, on t*, you *recognize* the person you are seeing is the person you saw on t. Your t-perception is given by existential information, and so is your t*-perception. So if, on t*, you *recognize* that you are seeing the same person you saw on t, that must consist in your realizing that the same entity satisfies both existence-claims are satisfied – that there is some one entity who satisfies both. Once a few platitudes are granted about sense-perception – that sense-perceptions have existential content -- our theory of conception and recognition proves to be far from *ad hoc.*

I should now deal with what is probably the most important objection of all to our analysis: the objection that is epistemological underpinnings are bankrupt.

You say that (k) and (i) are synonymous but differ in cognitive value. That is defensible, probably even true. But your explanation of *why* they differ in cognitive value is dubious at best. It seems to involve a false epistemology: one that makes our awareness of things less direct than it is.

Let us review your analysis. You say that one makes one's way from a token of "Smith" to O by way of some existence claim – one that corresponds to the peculiarities of the circumstances under which one encounters that expression. You make a corresponding claim about "Bob". Starting from

these assumptions, you arrive at the conclusion that what “Bob is a great racquetball player” communicates *to you* is some existence claim, and that what “Smith is a great racquetball player” is some other existence claim. In this way, you explain the cognitive difference between those two sentences, while conceding their synonymy.

But this story is inconsistent with the fact that we are aware of *objects.* When you first meet O (Bob/Smith), you are told “This is Bob”. Under those circumstances, your learning the meaning of “Bob” involves your seeing (or otherwise perceiving) Bob. So *Bob* himself is an object of your visual perception. And you learn that *that* entity, the one we just mentioned, is the semantic content of “Bob”. In the story you told, your learning what “Smith” means involves your having a perception of *that* entity. You see Smith; that very entity – O/Smith/Bob -- is an object of your awareness. And when you are told “that is Smith”, you are learning that *that very entity* – O/Smith/Bob -- is called “Smith”. No existential propositions serve as intermediaries: you are, in both cases, aware of Smith/Bob/O. In the one case, you learn that *that* entity is called “Smith”; in the other you learn that *that* entity is called “Bob”. And *that* entity is safely ensconced in your awareness: you don’t need to make your way to it via any existential, or other, propositions. So, it seems to me, that the epistemological conceit underlying your analysis is quite false. None of these mediating existential propositions exist; our cognitive access to objects, and therewith to semantic content, is much more direct than your Berkeleyan epistemology suggests.

If the story told by the objector were true, it is hard to see how “Bob is a great racquetball player” and “Smith is a great player” (and “Hesperus is lovely” and “Phosphorous is lovely”…) could possibly differ in cognitive

value. For, if the objector is right, our awareness of objects is so immediate and unremitting, and our ability to access semantic content so uninhibited and unconditional, that one could scarcely have any idea what "Hesperus" and "Phosphorous" meant without *ipso facto* knowing them to be co-referential.

Independently of that, the objector's point is based on a spurious epistemology. Obviously we can be aware of objects - of dogs, planets, and so forth. Barry Stroud (2001) made the profound point that perception is necessarily "predicational". You don't *just* see Fido; you see Fido *as* having these or those properties. In seeing an object, you necessarily "predicate" certain properties of it. I submit that once this is granted, the way is cleared for the analysis given above. Permit me to defend and delineate this view.

We see objects: that is given. You see Fido: that is a given. But you don't *just* see Fido. You see *a* thing that occupies a certain part of space, and time; that has floppy ears and a wet nose; that appears to under-nourished or over-nourished; that is chasing a certain other object…What is given to you is existential information: *there is an entity occupying a certain location, moving about certain ways, chasing a certain other entity…*

A perception that was *just* a perception of Fido would be a surd. Whenever Fido is given to you, he is embedded, so to speak, in existential information like that just described. That existential content is obviously not incidental to your perceptual awareness of Fido: there is no way to get rid of it without getting rid of the perception of Fido. Given this, it is not entirely easy to see how that existential content could fail to *constitute* the content of your perception of Fido. And if, under the circumstances described, you were told, while having these perceptions, "*that* is Fido" or (more pedantically) "our label for that thing is 'Fido'", then that content would figure into your subsequent interpretations of tokens of "Fido", presumably in the manner described above. Thus, far from being absurd, the epistemological conceit

underlying my analysis seems *de rigueur*. If we deny it, we come perilously close to saying – in fact, we *do* say – that one can *just* see Fido, without seeing *a creature, in a certain place, with such and such other properties…* But one cannot see an object – an object cannot be perceptually given to one – without one's being given all of this existential content.

Of course, this thesis about perception is not an innocuous one. Some will perhaps deny it.[359] But, as epistemological theses go, it is a pretty hearty one. And once it is granted, not much else is needed to generate the analysis given above concerning the cognitive differences between "Smith is a great racquetball player" and "Bob is a great racquetball player".

Bach, Kent. 1987. *Thought and Reference.* Oxford: Clarendon Press.

Kaplan, David. 1986. "Opacity". In *The Philosophy of W.V.O. Quine.* Paul Schilpp and Lewis Hahn ed.. La Salle, Ill: Open Court Press.
--1990/1968. "Quantifying In". In *The Philosophy of Language.* A.P. Martinich ed.. Oxford: Oxford University Press.

Kripke, Saul. 1972. *Naming and Necessity.* Cambridge: Harvard University Press.
1987/1979. "A Puzzle about Belief". In *Propositions and Attitudes.* N. Salmon and S. Soames ed.. Oxford: Oxford University Press.

MacCawley, James. 1983. *Everything Linguists Always Wanted to Know about Logic.* Cambridge: The M.I.T. Press.

Moore, Robert C. 1995. *Logic and Representation.* Palo Alto: CSLI Publications.

Neale, Stephen. 1990. *Descriptions.* Cambridge: The M.I.T. Press.

Perry, John. 2000. *The Problem of the Essential Indexical and Other Essays.* Palo Alto: CSLI Publications.

Quine, Willard van Orman. 1960. *Word and Object.* Cambridge: The M.I.T. Press.

Russell, Bertrand. 1990/1905. “On Denoting”. In *The Philosophy of Language.* A.P. Martinich ed.. Oxford: Oxford University Press.

Strawson, P.F. 1990/1950. “On Referring”. In *The Philosophy of Language.* A.P. Martinich ed.. Oxford: Oxford University Press.

Knowledge of Referent versus of Knowledge of Meaning: Another reason to deny that sentences refer to truth-values

There is another reason to think that sentence-tokens don’t refer to truth-values. You can obviously have a complete knowledge of what a sentence-tokens means without knowing whether it is true or not. So if sentence-tokens refer to truth-values, then it follows that, in at least some cases, one can understand a referring expression without knowing what it refers to. But, I

submit, this is not the case: under no circumstances can you possibly know what a referring term means without knowing what it refers to. Let “Blark” be an expression that refers to somebody, but suppose you don’t know who it refers to. In that case, you don’t know what is meant by “Blark snores”, “Blark is tall”, and the like. The reason, presumably, is that you don’t know what “Blark” means – you know what all the other component expressions of those sentence-tokens mean. And the reason you don’t know what “Blark” means is, presumably, that you don’t know to whom it refers.

There is an obvious class of apparent counter-examples to our thesis. Definite descriptions seem to be referring terms, and surely we can understand such expressions without knowing what they refer to. I know what is meant by “the man who murdered Smith” without having any idea who murdered Smith. This is reflected in the fact that, without knowing who murdered Smith, I fully understand tokens of sentences like “the man who murdered Smith is rich”, “the man who murdered Smith is bald”, and so on. I know what is meant by “the number of lines in the shortest version of the most important mathematical proof of the 24th century” without having any idea what that number is. This is reflected in the fact that, without knowing what that number is, I know what is meant by sentence-tokens like “the number of lines in the shortest version of the most important mathematical proof of the 24th century is less than ten-thousand”.

Here is why I think that, despite all this, our thesis prevails. Suppose that Fred is the man who murdered Smith. Given this, consider the sentence-token:

(*) “The man who murdered Smith is rich”.

As we noted, there are two ways to read (*): referentially or attributively. In other words, the definite description can be read as a referring term or as a

quantifier. Read referentially, (*) means:

(*R) *Fred is rich.*

Read attributively, (*) means:

(*A) *Somebody murdered Smith, and that person (whoever it is) is rich.*

If you don't know who murdered Smith, you cannot give a referential reading to (*): you are forced to give it an attributive reading. Suppose you have no idea who murdered Smith: somebody comes up to you and says (*). There is no doubt that you can have *some* kind of understanding of what that person is saying. But what is it that you understand? What meaning can you assign to that person's utterance? Surely not (*R). No – what you take away from that communiqué is (*A). Suppose somebody who you know to be absolutely reliable says to you: "the man who murdered Smith is rich". What do you now know? You surely don't know that *Fred* is rich or that *Mary* is rich. What you now know is *not* anything of the form: *O is rich*. What you now know is that *somebody* or other murdered Smith, and that, whoever that person might be, he or she is rich. Given that you don't know *who* murdered Smith, you are forced to interpret (*) *attributively.*

There is no doubt that, on *some* legitimate reading of it, you are able to understand (*) without knowing who murdered Smith. But you understand (*) in the sense that you can interpret it *attributively.* So you understand (*) in the sense that you know what meaning to assign it in the case where the definite description is operating as a *quantifier*, not as a singular term. In so far as (*) is operating *referentially*, you don't know what meaning to assign it – you don't know what it means. And this is obviously because, when (*) is operating

referentially, you don't know what the definite description refers to. So, yes, it is true that you can understand "the person who murdered Smith" without knowing what it refers to: but only in so far as that expression is amenable to an *attributive* reading – only in so far as it can be read as a quantifier, as opposed to a singular term. Considered as a device of reference, you know what that expression means only to the extent that you know what it refers to.

This point generalizes to all putative counter-instances of our thesis. Suppose (what is presumably true) that you have no idea what number of lines is in the shortest version the most important mathematical proof of the 24th century. There is no doubt that, under that circumstance, you can have *some* kind of understanding of:

(**) "the number of lines in the shortest version of the most important mathematical proof of the 24th century is less than ten-thousand".

But, under that circumstances, just what kind of meaning *can* you assign to (**)? You surely cannot take it to mean:

(**R) *twelve is less than ten-thousand.*

You don't know whether it refers to twelve or a thousand. Since you don't know what referent to assign it, you cannot take (**) to mean (**R). For any specific number n, you cannot take (**) to mean that *n* is less than a thousand. You don't know what the definite description refers to. No – so far as you can understand (**), it is because you can assign some entirely *general* proposition to it. Suppose somebody whom you know to be utterly reliable comes up to you and says (**). What do you now know? Not that *twelve* is less than a thousand, not that eighty is less than a thousand. What you know is some

purely general proposition:

(**A) There is *some* number n such that n numbers the lines in the shortest version of the most important mathematical proof of the 24th century, and n is less than ten-thousand.

So far as you can understand (**), it is because you can give it the right *attributive* reading. You know what (**) means in so far as it is can be taken to mean (**A). So in so far as you can understand that sentence-token, it is because the definite description is operating as a quantifier, not a singular term. In so far as (**) means (**R), you simply *don't* know what (**) means; and that is obviously because, so far as (**) is to be read that way, you don't know what the definite description refers to. So when the definite description is operating referentially, i.e. as a referring term, you *don't* understand (**); and that is obviously because you don't know how to interpret the definite description.

Let us sum up. To the extent that you *do* know how to interpret the definite description, i.e. to the extent that you do know what it means, it is to be read attributively, and thus as a quantifier. To the extent that it is to be read referentially, i.e. as a referring term, you *don't* know what it means, given that you don't know what the number of lines is in the relevant proof. So in so far you can understand "the number of lines in the shortest version of the most important mathematical proof of the 24th century" without knowing to what that expression refers, it is because that expression is *not* a referring term. Given *any* definite description, it can be shown, by an argument exactly analogous to that just given, that *either* there is no understanding that expression without having a knowledge of its referent, *or* that expression is not a referring term at all. No doubt I have *some* understanding of what is meant

by

(***) “the chemical compound that cures AIDS contains gold”,

even though I don’t know what the relevant chemical compound is. But so far as I know what is meant by (***), it means:

(***A) there is *some* chemical compound x such that x cures AIDS; moreover, x contains gold.

Given any specific chemical compound XYZ, I cannot take (***) to mean:

(***R) XYZ contains gold.

So far as (***) is to be read referentially, I do not, at this point in medical history, know what it means. And this is obviously because, considered as a device of reference, I don’t know how to interpret the definite description. In so far as I *do* know how to interpret it, the definite description is to be read attributively, and thus as a quantifier.

To sum up, definite descriptions are *not* counter-examples to our thesis that there is no knowledge of what a referring term means absent knowledge of what it refers to. On the contrary, that thesis is *vindicated* by consideration of definite descriptions. For in so far as “the chemical compound that cures AIDS” can be understood, absent a knowledge of the identity of the relevant compound, that expression is *not* an expression that refers to that compound. Absent a knowledge of what it refers to, “the phi” can be understood only in so far as it is *not* an expression that refers to the unique phi.

Now we can close the argument. Consider some sentence-token whose

truth-value you don’t know:

(B) “George W. Bush smoked a cigarette on July 21, 1977”.

You have a complete understanding of that expression. There is *no* sense in which your ability to interpret that expression is compromised by the fact that you don’t know its truth-value. Understanding a sentence-token is obviously a pre-requisite to knowing its truth-value. So knowledge of truth-value cannot possibly a pre-requisite to understanding. Also, there is no clear sense in which (B) is ambiguous between referential and attributive readings. (B) is obviously unambiguous, and is not remotely like “the person who murdered Smith is rich”. We have seen that, notwithstanding a few apparent counter-examples, there can be no complete understanding of a referring term absent a knowledge of what it refers to. There obviously is, and must be, a complete understanding of sentence-tokens absent knowledge of their truth-values. (It would make no sense to say that, in order to have a full understanding of what a sentence-token means, one must first know whether it is true or not.) Therefore sentence-tokens do not refer to truth-values.

Quantification in Natural Language

1.0. The five contentions that it is the purpose of this paper to establish

The present paper will defend the following five points:

Point #1: “Some phi is a psi” is *not* an existence-claim; it is *not* to the effect that there exists at least one phi and, moreover, at least one phi is also a psi.

Point #2: “Some phi is a psi” merely *presupposes* that there is at least one phi. It does not (see Point #1) *say* it. (By contrast, “there exists at least one phi and, moreover, at least one phi is also a psi” actually *says* it.)

Point #3: The sentence “all phi’s are psi’s” is *incompatible* with the sentence “some phi’s are psi’s.” So the former entails the *negation* of the latter and *a fortiori* isn’t compatible with the latter.

Point #4: “Five elephants can dance” is to the effect that exactly five elephants can dance. So that sentence is false, and not merely misleading, if six elephants can dance. For any numeral n, “n phi’s are psi’s” is to the effect that exactly n phi’s are psi’s.

Point #5 (an immediate consequence of Point #4): “Ten elephants can dance” is incompatible with the sentence “five elephants can dance.” In general, if $m \neq n$, “m phi’s are psi’s” is incompatible with “n phi’s are psi’s.”

Most contemporary logicians and philosophers either assume, or explicitly state, that each of #1-#5 is false, and much of what they say crumbles unless they’re right about this.

Section 2.0. Why Point #1 is correct But #1-#5 are correct. Right now let’s focus on #1.

Consider the following two sentences:

1. “There is at least one diamond on the table, and at least one such diamond is extremely valuable.”

2. “Some of the diamonds on the table are extremely valuable.”

For argument’s sake, assume that there are several diamonds on the table in question and that many of them, if not all, are extremely valuable.

Do 1 and 2 have the same meaning? No. 1 explicitly says, among other things, that there is at least one diamond on the table. But 2 doesn’t *say* this; it *presupposes* it. If it is in question whether there are any diamonds on the table, an utterance of 1 *would* be appropriate, whereas an utterance of 2 would *not* be appropriate. Why is it inappropriate to utter 2 under such circumstance? Because, clearly, such an utterance presupposes an answer to a question that, given such circumstances, has yet to be answered. Why is it appropriate to utter 1 under such circumstance? Because, clearly, such an utterance *gives* answer to a question that that, given such circumstances, has yet to be answered. So, contrary to what is generally held, 2, unlike 1, is *not* an existence-claim.

2.1 A similar argument If asked “do any felines exist?”, one is responding in a completely non-defective manner if one says:

(A) “Yes. There exists at least one feline; moreover, at least one feline is ferocious.”

But one *is* responding in a defective manner if one instead says:

(B) “some felines are ferocious.”

Under the circumstances, what is in question is whether there are any felines. B presupposes an answer to that very question; that’s why it’s inappropriate. A does not presuppose an answer to that question; A *asserts* that there is at least one feline: unlike B, it doesn’t just *assume* it. And that’s why, under the circumstances, A *is* an appropriate utterance.

Thus, whereas it is (part of) A’s job to state that there are felines, it is no part of B’s job to do this. This shows, in other words, that B is an existence-claim, whereas A is not. Points #1 and #2 have thus been established.

Notice also that, if asked “do any felines exist?”, one is *over*responding by uttering A, whereas, if asked that same question, one is *under*responding by uttering B. And if, having been asked that question, one were to utter A, one’s interrogator would *not* respond with puzzlement, whereas one’s interrogator *would* so respond if one had uttered B. Why this difference? The obvious answer is this: whereas B isn’t *quite* on point, A is on point. Why *this* difference? In other words, why is it that B, unlike A, isn’t quite on point? The obvious answer is this: in uttering B, one is presupposing what, given the question, one should be affirming, whereas, in uttering A, one is not making that mistake.

2.2 Presupposition vs. assertion The following story will help clarify how A and B are semantically different. If, while pointing at a certain man, I utter the sentence:

(MS) “that person is a menace to society”,

what I am *saying* (in other words, the proposition I'm affirming) is not to the same effect as what I'd be saying if *ceteris paribus* I were to utter the sentence:

(TE) "there exists a uniquely contextually salient individual x; moreover, x is a menace to society."

In uttering TE, I am asserting, not presupposing, that there is such a man. And that's why, if asked whether there a contextually salient man, I'd be responding in a non-defective manner by uttering TE. In uttering MS, on the other hand, I am presupposing, not asserting, that there exists a uniquely contextually salient person x. And that's why, if asked that same question, I'd be responding in a defective manner by uttering MS.

Bearing in mind what we just said, let's revisit sentences A and B. As we saw, A is, whereas B is not, an appropriate response to the question "do any felines exist?" And the reason, as we said, is that, whereas A is itself an answer to that question, B assumes a certain answer to it. Hopefully, the points just made in connection with ME and TE reinforce what we said about A and B and, therefore, about 1 and 2.

Section 2.3: The concept of presupposition and the distinction between sentence-logic vs. proposition-logic "There must be an error in your reasoning", it will be said:

"1 is obviously an existence-claim. You admit this. You also admit that 1 and 2 are equivalent. Given those facts, it surely follows that the proposition affirmed by 2 is equivalent with one to the effect that there exists at least one apple on the table and, moreover, that at least one such apple is rotten. And

once *that* is granted, it becomes very hard to see how 2 could fail to be an existence-claim. In fact, given that 1 and 2 are equivalent, it becomes unclear what it even means to say that 2 is not an existence-claim."

1 and 2 are indeed equivalent. (In any case, they're equivalent in the weak sense that the one cannot be true if the other fails to be true, even though, in circumstances where there is no unique contextually salient man, TE one is false whereas ME other is without truth-value.) But 1 and 2 are *sentences*. Sometimes equivalent *sentences* have non-equivalent *propositions* for their meanings. This is one of those cases. Once again, it will help to look at TE and MS. The semantic rule for MS is this:

(MSR) If, in a given context C, there is a uniquely contextually salient individual x, then the proposition meant by an utterance in C of "that person is a menace to society" is identical (or equivalent) with: *x is a menace to society*.

The semantic rule for TE is this:

(TER) The proposition meant by an utterance, in a given context C, of "there exists a uniquely contextually salient individual x; moreover, x is a menace to society" is identical (or equivalent) with: *there is a unique contextually salient individual x and, moreover, x is a menace to society*.

Thus, supposing that, while pointing at Smith, I utter ME, the proposition I'm thereby affirming is true iff Smith is a menace to society. That proposition has nothing to do with Smith's being contextually salient; that proposition would be true even if Smith weren't contextually salient. The material

relating to contextual salience in the semantic rule for ME does not itself make it into the proposition meant by a given utterance of that sentence; that material serves the purpose only of *identifying* the person whom such an utterance (when non-abortive) describes as being a menace to society.

But the proposition I'm affirming in uttering TE---even if, while doing so, I'm pointing at Smith---could be true even if Smith *weren't* a menace to society; and that proposition *does* contain material relating to the concept of contextual salience (for it says *inter alia* that something x falls under that concept).

At the same time, the semantic rules for MS and TE being what they are, there is no possible circumstance in which an utterance of the one sentence is true and in which an utterance of the other is false. (By "no possible circumstance", I mean "no possible circumstance *in which* the semantic rules of the English language are operative.") So, even though the proposition literally meant, in a given set of circumstances, by a given token of ME is not equivalent with the proposition literally meant, in those very circumstances, by an utterance of TE, there is no possible set of circumstances in which an utterance of MS isn't true exactly iff an utterance of TE is true.

A consequence of the reasonable presumption that MSR is the semantic rule for MS is that, if produced in circumstances in which there is no contextually salient individual x, an utterance of MS fails to have *any* proposition for its meaning and is therefore abortive (i.e. neither true nor false). MSR is to the effect that, *given* such an individual, a person who utters MS is thereby describing that individual as menace to society; MSR makes no provision for circumstances in which there is no such individual. It is for that reason that if, under such circumstances, one were to utter MS, that statement would register, not as false, but as out of place. And it is for that reason *mutatis mutandis* that, were one to utter TS under such circumstances,

that statement would register as false, and not as out of place.

3.0 Why Point #3 is correct The orthodoxy these days is that (in a non-empty universe) the sentence:

(SP) “something is a phi”

is a logical consequence of the sentence:

(EP) “everything is a phi.”

(From now on, we’ll leave the parenthetical qualification unstated.)

Let SE be the contention that SP entails EP. (Of course, SP is compatible with EP if it entails it.) SE concerns sentences of ordinary English; it doesn’t concern their artificial analogues. So SE is *not* the principle, which we’ll call “SEF”, that

(EPF) “(x)(x has phi).”

entails

(SPF) “(Ex)(x has phi).”

SEF is trivially correct, given that EPF and SPF are *stipulated* to mean, respectively,

(EPF#) given any object x, x has phi

and

(SPF#) there exists at least one thing x such that x has phi.

SE says that SP is a consequence of EP. SE doesn't say anything about SPF# or EPF#.

3.1 Why Point #3 is correct: a preliminary argument Is SE correct? No. A story will help make this clear. You go to a costume party. Every single person at the costume party is wearing a vampire costume. Later, when you are talking with your roommate Smith, who wanted to attend the party but was unable to, Smith asks you: "how many people at the party wore vampire costumes?" You answer by saying:

(SVC) "some of the people at the party wore vampire costumes."

Surely you've given the wrong answer. What you should have said is:

(AVC) "*all* of the people at the party wore vampire costumes."

The fact that AVC is, whereas SVC is not, the right answer does not, by itself, prove that SE is false. That fact suggests, in other words, that Point #3 is correct.

3.2 An objection to an objection to our preliminary argument It has been argued that SVC is technically correct (i.e. that the proposition that is its literal meaning is correct), even though, given the circumstances, SVC misleadingly *suggests* that some people did not wear vampire costumes.

But this position is demonstrably false. Consider the following sentences:

(a_1) Some cupcakes are yummy.
(a_2) Only some cupcakes are yummy.
(a_2*) Only *some* cupcakes are yummy. (The italics indicate that the occurrence of "some" is being emphasized.)
(a_3) at most some cupcakes are yummy.

It should be noted that a_2 and a_3 are (virtually) synonymous and, in addition, that each of a_1-a_3 is perfectly grammatical and perfectly meaningful.

Bearing these points in mind, let us suppose for argument's sake that:

(S) "some phi is psi" is to the same effect as "at least one phi is psi."

If S is true, then

(a_2) "Only some cupcakes are yummy"

is to the effect as:

(a_2?) *only* the members of a proper subset of the set of cupcakes *or* of the entire set of cupcakes in its entirety are yummy.

So far as a_2? has a clear meaning, it seems to be to the effect that:

(a_2?*) at the very least, the following holds: either every member of some

proper subset of the set of all cupcakes are yummy or every member of the set of cupcakes (i.e. of an *improper* subset of the set of cupcakes) are yummy.

But this is obviously *not* what a_2 means. The occurrence of "only" in a_2 tells the auditor to take the word "some" at face value; it tells him, in other words, to *interpret it literally*.

3.2.1 The semantics of "literalizing" terms, such as "only", "at most", and "literally" A story will help clarify some of the points just made. While driving together, you and I have the following dialogue:

You: Is there a nearby place to get food?

Me: There's a McDonalds down the road.

You: Given what you said, it follows that we will soon have access to food, the reason being that food is available at the nearby McDonalds."

Me: No—that doesn't follow. Given *only* what I said, it could be that the McDonald's is closed; or it could be that we don't have any money, and therefore can't get food there, even if it's open.

You: I grant that the *only* thing you *said*---i.e. that the *only* thing that, at the level of literal meaning, you affirmed---was that there was a McDonald's down the road. But, under the circumstances, I was right to make the inference that it wasn't *only* the case that there was a McDonald's down the road; I was entitled to make the inference that said McDonald's would be a source of food.

Me: Yes, you're right. But so am I. For, as you yourself grant, I said *only* that there was a McDonald's down the road; and, as you yourself grant, given only what my words *literally* meant, I left it open whether or not said McDonald's would be a place we could get food.

In this story, I'm guilty of violating the usual conversational protocols and, in so doing, I'm being extremely obstreperous. But from a strictly semantic perspective, my end of the conversation is quite impeccable.

In this conversation, the occurrences of the word "only" are very clearly there to direct you, the auditor, to confine your attention to the *literal meaning* of what I said, and not to read non-literal meanings into what I said.

It is in precisely this way that the occurrence of "only" in a_2 is functioning. In uttering a_2*, one is unambiguously saying that all and only the members of a *proper* subset of the set of all cupcakes are yummy.

Moreover, it would be bad methodology to deny that this was the literal meaning of a_2. For such a denial would be tantamount to a denial of the fact that our intuitions (though clearly defeasible, especially in matters relating to semantics) are to be given *some* weight.

So we must conclude that, at the level of literal meaning, a_2 is to the effect that all and *only* the members of some *proper* subset of the set of all cupcakes are yummy. An exactly similar argument holds with respect to a_3 and thus establishes the very same conclusion.

3.2.2: Why, though semantically redundant, literalizers (e.g. "only", "literally") are not redundant *tout court* "But", it might be asked, "doesn't your analysis entail that the occurrence of 'only' in a_2 is redundant and, therefore, that it is serving no function?"

My analysis may entail that the aforementioned occurrence is *semantically* redundant. But that doesn't mean that it is redundant *tout court.* (Sometimes what is redundant from the viewpoint of semantics is non-redundant from the viewpoint of pragmatics. This is one of those times.)

Many expressions are such that their semantic function is to confine the auditor's attention to literal meaning---are such their function is to tell the auditor not to draw inferences that go beyond what, at the level of literal meaning, was affirmed. Among such expression are: "exactly", "precisely", and, indeed, "literally."

The following dialogues will make this clear:

Dialogue #1

Smith: How much money does that guy have?
Jones: He has a billion dollars.
Smith: No, really, how much money does he have?
Jones: He *literally* has a billion dollars.

Commentary: Clearly, Jones' utterance of "[h]e *literally* has a billion dollars" is to the effect that his previous utterance of "[h]e has a billion dollars" is to be taken literally.

Dialogue #2:

Smith: How much money does Brown owe to the bank?
Jones: Brown owes $10,000 to the bank.
Smith: I need to know *precisely* how much money Brown owes the bank.
Jones: Brown owes *precisely* $10,000 to the bank.

Commentary: Clearly, Jones' utterance of Brown owes *precisely* $10,000 to the bank" is to the effect that his previous utterance of Brown owes $10,000 to the bank." is to be taken literally.

Dialogue #3:

Smith: When were human beings first able to talk?
Jones: Human beings were first able to talk 100,000 years ago.
Smith: I need to know *exactly* how when were human beings first able to talk.
Jones: Human beings were first able to talk *exactly* 100,000 years ago. In other words, whatever today's date is, and whatever the exact moment was at which I uttered the previous sentence, subtract 100,000 from that, and that will give you the time—the exact time---when people first spoke.

Commentary: The same as last time, *mutatis mutandis.*

A conversation of any length or depth that consisted of statements all of which were to be taken literally would be very different from any actually conversation. Much of what is communicated is not what is literally meant by our words, and is instead (non-literally) suggested. (It is through "conversational implicatures", to use H.P. Grice's now standard term, that we do much of our communicating.) But sometimes we want our words to be interpreted narrowly; we don't want people to read extra-semantic meanings into them. It should therefore be no surprise that language contains devices whose very purpose is to make it clear that the utterances in whose scope they fall are to be taken literally.

4.0 Why points #4 and #5 are correct These same dialogues make it clear that numerals mean exactly what they seem to mean. (They make it clear that "10,000" means *10,000*: not *at least 10,000* or *approximately 10,000.*) And the way in which these dialogues makes this clear is very similar to the way in which they make it clear that "some" means *some,* as opposed to *at least some (and possibly all)*). Here we need only repeat what we said a moment ago, making a few adjustments along the way.

Consider the following sentences:

(b_1) "Brown owes $10,000 to the bank."
(b_2) "Brown owes *precisely* $10,000 to the bank."
(b_3) "Brown owes *at most* $10,000 to the bank."
(b_4) "Brown owes at least $10,000 (i.e. he either owes exactly that amount or he owes more)."

For argument's sake, suppose that "10,000" means "at least 10,000", and make the correlative supposition that the literal meaning of b_1 is to the same effect as the literal meaning of b_2. In that case, b_2 is to the same effect as the sentence (or, in any case, sentence-like expression):

(b_2?) "Brown owes *precisely* at least $10,000 to the bank."

So far as b_2? has a clear meaning, it seems to be to the effect that:

(b_2?*) the following is unqualifiedly (literally, technically, etc.) correct: *either* Brown owes $10,000 to the bank (and not a penny less) *or* he owes more than $10,000 to the bank.

But b_2?* clearly isn't what b_2 means. b_2 says that Brown owes $10,000 to the bank: not one penny more and not one penny less. An exactly similar argument holds with respect to b_4 and thus establishes the very same conclusion, viz. that "10,000" means *10,000*: not *at least 10,000* or *roughly 10,000*.

Duly generalized, this argument establishes that, for any numeral n, n means *n*: not *at least n* or *roughly* n. Thus, Point #4 is correct; and Point #5 is also correct, given that #5 is obviously an immediate consequence of #4.

4.1 A similar argument Consider the following argument/sentence:

(^) Given that Smith has seven cars, he has an odd number of cars.

If "seven" meant *at least seven*, then ^ would not express a valid argument. But it clearly does express a valid argument.

In response to this, one could take the position that, at the level of literal meaning, ^ does not convey a valid argument and that, to the extent that tokens of it *communicate* such an argument, it is by the grace of implicatures. But there is no independent support for such a position, which is an implausible one and for the negation of which we'll now provide additional support.

In the sentence

(*) "5+5=10"

the number-expressions (numerals) are, or at least appear to be, functioning as substantives (nouns). (The occurrences of "5" and "10" seem to be synonymous with, respectively, "the number 5" and "the number 10"; and, of

course, those definite descriptions are noun-phrases.)

But in the sentence

(**) “5 things plus 5 other things add up to *exactly* 10 things.”

those same numerals are functioning as adjectives.

That said, * and ** have extremely similar meanings. But if the occurrences in ** of “5” meant *at least five*, then, even though * is true, ** would be false, a consequence being * and ** had different meanings. * is indisputably to the effect that 5+5=*exactly* 10; but, given its similarity in respect of meaning to **, * would NOT be to that effect IF “5” meant “at least five.”

A similar argument: consider the sentence:

(***) If Jim has five cars in Idaho, and he has five cars in Florida, and he has no other cars besides those just mentioned, then, Jim has exactly ten cars altogether.

*** is correct. (We’re assuming that none of Jim’s Idaho-cars is numerically identical with any of his Florida-cars.) But if the occurrences in *** of “5” meant *at least* five, *** would be false.

So “5” means what it seems to mean. It means *five*: not *at least five* or *roughly five*. The same line of reasoning *mutatis mutandis* holds for any given numeral n. Thus, point #5 is correct. We have thus established the truth of Points #1-#5.

References

Chomsky, Noam. 1988. Language and the problems of knowledge. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 509-527.

Donnellan, Keith. 1966. Reference and definite descriptions. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 235-247.

Dummett, Michael. 1973. Frege: Philosophy of Language. Harvard University Press, Cambridge, MA.

Evans, Gareth. The Varieties of Reference. Clarendon Press, Oxford.

Frege, Gottlob. 1892. On sense and nominatum. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 190-202.

Jackendoff, Ray. 1989. What is a concept, that people may grasp it? In: Laurence, E., et al. (Eds.), Concepts: Core Readings. The M.I.T. Press, Cambridge, MA, pp. 305-334.

Kaplan, David , 1970. Dthat. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 316-329.

Kaplan, David. 1989. Demonstratives. In: Almog, J., et al.(Eds.), Themes from Kaplan. Oxford University Press, New York, pp. 471-555.

Kaplan, David. 1989b. Afterthoughts. In: Almog, J., et al. (Eds.), Themes from Kaplan. Oxford University Press, New York, pp. 557-534.

Kripke, Saul. 1968. Speaker's Reference and Semantic Reference. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 248-258.

Kripke, Saul. 1977. A Puzzle about Belief. In: Margalit, A. (Ed.), Meaning and Use. D. Reidel, Dordrecht.

Neale, Stephen. 1990. Descriptions. The M.I.T. Press, Cambridge, MA.

Quine, Willard van Orman. 1953. Reference and Modality. In: Quine, W.V.O. (Ed.), From a Logical Point of View. Harvard University Press, Cambridge, MA, pp. 139-151.

Quine, Willard van Orman. 1960. Word and Object. The M.I.T. Press, Cambridge, MA.

Recanati, Francois. 1989. The Pragmatics of What is Said. Mind and Language 4, 295-329.

Recanati, Francois. 1993. Direct Reference: From Language to Thought. Oxford: Blackwell.

Recanati, Francois. 2004. Literal Meaning. Cambridge: Cambridge University Press.

Russell, Bertrand. 1905. On Denoting. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 203-211.

Russell, Bertrand. 1920. Introduction to Mathematical Philosophy. London: George Allen & Unwinn.

Salmon, Nathan. 1986. Frege's Puzzle. Cambridge: M.I.T. Press.

Salmon, Nathan. 2007. Content, Cognition, and Communication: Philosophical Papers II. New York: Oxford University Press.

Soames, Scott. 2002. Beyond Rigidity: The Unfinished Semantic Agenda of *Naming and Necessity*. New York: Oxford University Press.

Sommers, Fred. 1982. The Logic of Natural Language. New York: Oxford University Press.

Strawson, Peter. 1950. On Referring. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 219-234.

Intensionality

Abstract: On the basis of arguments put forth by Kripke (1970, 1980), it is widely held that one can sometimes rationally accept propositions of the form *P and not-P* and also that there are necessary *a posteriori* truths. We will find that Kripke's arguments for these views appear probative only so long as one fails to distinguish between semantics and presemantics – between the literal

meanings of sentences, on the one hand, and the information on the basis of which one identifies those literal meanings, on the other. This same failure, it will be argued, underlies the popular thesis that intersubstituting coreferring terms sometimes turns true sentences into false ones and *vice versa*. Though seemingly plausible, this thesis has a number of counterintuitive consequences, among them that the occurrence of "snow" in "it is true that snow is white" doesn't refer to snow. An understanding of the distinction between semantics and presemantics suggests a way to develop a semantic system that doesn't have these consequences and that, moreover, reconciles our intuitions concerning cognitive content with some powerfully argued theses of contemporary philosophy of language. Some of this paper's main contentions are anticipated by Andrzej Boguslawski in his 1994 paper "Sentential Complementation and Truth."

I. Introduction

In at least some contexts, replacing a referring term with a co-referring term preserves truth-value.

(1) "Hesperus is closer to the Sun than the Earth"

is true. "Phosphorous" and "Hesperus" refer to the same thing. So

(2) "Phosphorous is closer to the Sun than the Earth".

is also true.

If a context tolerates intersubstitutions of co-referring terms---in other words, if intersubstituting co-referring terms in that context cannot change

the truth-value of the host-sentence---that context is “extensional.” If an expression occurs in an extensional context, let us say that it occurs “extensionally.” “Hesperus” and “Phosphorous” occur extensionally in (1) and (2), respectively.

There seem to be cases where replacing a referring term with a co-referring term leads from truth to falsehood or *vice versa*. Given only that

(3) “John believes that Hesperus is closer to the sun than the Earth”

is true, it does *not* appear to follow that

(4) “John believes that Phosphorous is closer to the sun than the Earth”

is also true. It thus appears, at least at first, that not all contexts are extensional. If a context is not extensional, we will say that it is “intensional” and that the expressions occurring in it do so “intensionally.” “Hesperus” and “Phosphorous” appear to occur *intensionally* in (3) and (4), respectively.

Some striking apparent violations of extensionality involve intersubstitutions of definite descriptions. Given only that

(5) “Little Timmy believes that the whole number that comes immediately after one is less than three.”

is true, it very much seems not to follow that

(6) “Little Timmy believes that the number n such that n is a unique even prime if arithmetic is incomplete and is otherwise a unique positive square root of 71 is less than three.”

is also true.

Nonetheless, I will argue in this paper that *all* contexts are extensional and that, given a certain distinction, the reasons for thinking otherwise prove to be insubstantial. The distinction in question is that between, on the one hand, what is literally meant by an expression and, on the other hand, the information that one must work through to *assign* the right literal meaning to an expression that one encounters.

If P is the literal meaning of a sentence (or sentence-token) S, let us say that S "semantically encodes" P. Of course, what a sentence *communicates* may or may not coincide with what it semantically encodes. (An utterance of "sometimes one and one don't make two" may communicate a truth, even though what it literally means is false.) And, in this context, it is of the utmost importance that we bear that fact in mind. (In this paper, I will use "convey" as a synonym for "communicate.")

Russell's Theory of Descriptions is relevant to much of what we will say in this paper, and we should therefore make it clear at the outset what it is. The essence of that theory is given by the contention that:

(TD) If S is a sentence of the form "…the phi…", then in virtue of that fact S has for its meaning either the proposition *something O uniquely has phi and…O…*or some other, logically equivalent, quantified generalization.

Thus, according to Theory of Descriptions, "the king of France is bald" has for its meaning a quantified generalization that is either identical or equivalent with: *something x is uniquely a king of France; moreover, x is bald.* Thus, if that theory is correct, then "the king of France" is not a

referring term and is instead a quantifier (or, in any case, a quantifier-like expression).

II. Names

According to Frege (1892) and Russell (1905, 1918), "Socrates was wise" has for its meaning a proposition either identical or logically equivalent with one along the lines of: *something x was a unique philosopher of antiquity to die of hemlock poisoning; moreover, x was wise*. In their view, "Socrates" *describes* Socrates, and thus doesn't merely label him, the same thing *mutatis mutandis* being true of any other proper name.

In light of arguments put forth by Kripke (1980), there is overwhelming reason to believe that the Frege-Russell view is false. Supposing that the Frege-Russell view is in fact false, it is natural to suppose that proper names are mere labels, i.e. that names don't describe their referents and merely pick them out. If that view is correct, then there is some x such that a token of "Socrates was wise" has for its literal meaning the proposition *x was wise* and, more generally, such that the literal meaning of any sentence of the form "Socrates has phi" is simply *x has phi.* Thus, for some x, an utterance of "Socrates was wise" is true just in case x was wise, it being entirely irrelevant what other properties x, or anything else, happens to have or lack – it being entirely irrelevant, for example, whether x, or anything else, was a philosopher or a citizen of Athens or a drinker of hemlock. To put it in the most general possible terms: if proper names are mere labels, then given any proper name N whose referent is object O, the literal meaning of " N has phi" is simply *O has phi*. Thus, such an utterance is true exactly if O has phi, it being irrelevant what other properties O, or anything else, happens to have or lack. Many contemporary semanticists hold the view just described.

There are some apparently very compelling reasons to reject that view. “Hesperus” and “Phosphorous” are co-referring proper names. Supposing that those expressions are mere labels, there is some x such that both “Hesperus” and “Phosphorous” refer to x. In that case, “Hesperus has phi” has the same literal meaning as “Phosphorous has phi.” A corollary is that “Hesperus is identical with Phosphorous” has the same literal meaning as “Hesperus is identical with Hesperus.” But, as Frege (1892) pointed out, the one sentence communicates a significant astronomical discovery, while the other is an empty truism, and it is therefore very hard to see how “Hesperus” and “Phosphorous” could be mere labels. But, as previously noted, given the formidable arguments put forth by Kripke (1980), Soames (2002), and others, it is hard to see how they could not be mere labels.

Here is what we will see in this section. Supposing, if only for the sake of argument, that proper names are mere labels, there is no difficulty at all accounting for the dramatic communicative differences between, for example, “Hesperus is identical with Phosphorous” and “Hesperus is identical with Hesperus.” But those differences are to be explained *not* by supposing that those sentences differ in respect of literal meaning. Rather, those differences are to be explained in terms of facts about how people *ascertain* the literal meanings of the sentences that they encounter.

And we will find the following generalizations of this point to hold. First, for any predicate phi, supposing that “Hesperus” and “Phosphorous” are mere labels, there is no difficulty accounting for the fact that what is communicated by “Hesperus has phi” may differ dramatically from what is communicated by “ Phosphorous has phi.” Second, what was just said about “Hesperus” and “Phosphorous” is true *mutatis mutandis* of any two co-referring proper names. We will also see why, given the truth of these contentions, many (but not all) apparent threats to extensionality prove to be

illusory. (The remaining threats will be dealt with in subsequent sections.)

Let us begin. I see a man in the distance. I see that he has such and such properties (e.g. he is leaping about in a decidedly athletic way, he is wearing a bright-red shirt…). I am told that his last name is "Smith." I am also told:

(7) "Smith is a great racquetball player."

Given our supposition that proper names are mere labels, it follows that, for some object O, the proposition literally meant by (7) is:

(8) O is a great racquetball player.

The next day I meet somebody. He has thus and such perceptible properties (e.g. he is pudgy and unathletic looking). As I shake hands with him, I am told that his first name is "Bob." Although I don't know it, I am shaking hands with the man whom I saw earlier in the distance. In fact, let us suppose that, given the information at my disposal, I have no way of knowing that the man I saw yesterday is identical with the man with whom I am now shaking hands – that the laws of optics being what they are, my visual system being what it is, and so on, the information thus far given to me simply does not warrant my believing that I am now shaking hands with the man I saw yesterday.

With these points in place, let us continue our story. A moment after shaking hands with Bob, I am told

(9) "Bob is a great racquetball player."

Given that "Bob" is a mere label, and that it labels the same thing as

“Smith”, the proposition literally meant by (9) is (8).

At the same time, under the circumstances, (7) and (9) tell me different things. Indeed, what I learn in consequence of being told (7) does not entail, and isn’t entailed by, what I learn in consequence of being told (9).

“But,” it will be said, “this isn’t possible. (7) and (9) have the same literal meaning. Therefore what I learn from the one must coincide with what I learn from the other. So we are stuck with the paradoxical consequence that (7) and (9) do, and also do not, coincide in respect of what they tell me about the world.”

The fallacy that underlies this supposed paradox is easily exposed. What an occurrence of a sentence communicates to somebody isn’t always confined to its literal meaning (and, we will see, isn’t necessarily even inclusive of its literal meaning). To hear a sentence is not to hear a proposition. One must make one’s way from the physical noise that one hears to the proposition that it bears. When I hear (7), I have to make my way from the noise that I hear to the proposition that is that noise’s literal meaning. This involves, among other things, my assigning the right referent to “Smith.” To do this, I have to exploit background knowledge. In this context, the relevant background knowledge is given by some proposition at least approximately like the following:

(10) There is a person x such that x uniquely has such and such properties (viz. yesterday, from a considerable distance, I saw x leaping about in a decidedly athletic manner, and x was wearing a bright-red shirt…); moreover, “Smith” is x’s name.

At the level of semantics, “Smith” doesn’t describe Smith – it merely labels him. Nonetheless, in order for me to affix that label to the right person,

I must work through the just described background knowledge. That knowledge is existential in nature, its content being given by (10). A consequence is that the information through which I assign the right meaning to (7) is given by an existence-claim at least approximately like:

(11) There is a certain person x such that x uniquely has such and such properties (viz. yesterday, from a considerable distance, I saw x leaping about in a decidedly athletic manner, and x was wearing a bright-red shirt…), and such that "Smith" is x's name; moreover, x is a great racquetball player.

So even though (8) is what is literally meant by (7), the content of the knowledge that I must exploit to assign the right meaning to (7) is given by (11). It is thus no wonder that (7) *communicates* (11) to me, even though it has (8) for its literal meaning.

What we just said about (7) is true *mutatis mutandis* of any sentence of the form "Smith has phi." So if, on the day in question, I hear a token of "Smith is also a fine chess-player", what is communicated to me is some proposition at least approximately like:

(12) There is a certain person x such that x uniquely has such and such properties (viz. x is such that, yesterday, from a considerable distance, I saw x leaping about in a decidedly athletic manner and I saw that x was wearing a bright-red shirt…), and such that "Smith" is x's name; moreover, x is a fine chess-player.

Of course, as time passes, the existence-claim through which I access the meaning of (7) necessarily changes. Suppose that, two days after first seeing Smith, I am told: "Smith has decided to become a professional chess-player."

Given obvious extensions of the points already made, the existence-claim through which I assign the right literal meaning to that utterance is not:

(13) There is a certain person x such that x uniquely has such and such properties (viz. x is such that, yesterday, from a considerable distance, I saw x leaping about in a decidedly athletic manner and I saw that x was wearing a bright-red shirt…), and such that "Smith" is x's name; moreover, x has decided to become a professional chess-player.

Rather, it is:

(14) There is a certain person x such that x uniquely has such and such properties (viz. x is such that, two days ago, from a considerable distance, I saw x leaping about in a decidedly athletic manner and I saw that x was wearing a bright-red shirt…), and such that "Smith" is x's name; moreover, x has decided to become a professional chess-player.

As we will discuss later, there are other reasons why, with the passage of time, there must be changes in the identity of the information through which I must access the literal meaning of (7).

What we just said about (7) is true *mutatis mutandis* of (9). Hearing the noise "Bob" is one thing. Knowing what it refers to is another. To affix "Bob" to the right person, I must exploit background knowledge. Given the circumstances under which I am first told to whom "Bob" refers (see p. 4), the relevant background knowledge is given by a proposition at least approximately like the following:

(15) There is a person x such that x uniquely has thus and such characteristics

(viz. x is a pudgy and unathletic looking person with whom I shook hands a moment ago…) and “Bob” is x’s name.

Thus, even though there is some individual x such that, at the level of semantics, “Bob” merely labels x, and thus doesn’t describe it, the information *through which* I learn that “Bob” labels x *does* describe it.

Once again, suppose that, just after shaking hands with Bob, I am told:

(9) “Bob is a great racquetball player.”

In that case, for reasons analogous to those given in connection with (7), what is communicated to me will be some proposition along the lines of:

(16) There is a person x such that x uniquely has thus and such characteristics (viz. x is a pudgy and unathletic looking person with whom I shook hands a moment ago…) and “Bob” is x’s name; moreover, x is a great racquetball player.

By obvious extensions of this reasoning, what would be communicated to me, under the circumstances just described, by any sentence of the form “Bob has phi” would be:

(17) There is a person x such that x uniquely has thus and such characteristics (viz. x is a pudgy and unathletic looking person with whom I shook hands a moment ago…) and “Bob” is x’s name; moreover, x has phi.

Thus it is no mystery why, under these circumstances, (7) and (9) should

have entirely different cognitive values. (7) *communicates* (11), whereas (9) *communicates* (16). (11) and (16) are completely different propositions. Indeed, neither so much as entails the other. Thus, I can, without necessarily being irrational, assent to the one without assenting to the other.

Of course, what is semantically encoded in (7) *does* entail what is semantically encoded in (9), and *vice versa.* After all, they have the same literal meaning. But whether somebody is rational is to be understood in terms of what that person does with the information at his or her disposal. And, under the circumstances, the information that I have at my disposal in consequence of my being told (7) neither entails, nor is entailed by, the information that I have at my disposal in consequence of being told (9).

These points have an important bearing on the question of whether all contexts are extensional. A preliminary point is in order. Consider the sentence “snow is white.” The expression “that snow is white” refers to the proposition semantically encoded in “snow is white.” But, unlike “snow is white”, “that snow is white” is neither true nor false. The latter expression merely refers to a certain proposition, without taking a stand, so to speak, as to its truth or falsity. Just as “that snow is white” refers to the proposition semantically encoded in “snow is white”, so

(18) “that Smith is a great racquetball player”

refers to the proposition semantically encoded in:

(7) “Smith is a great racquetball player.”

And the expression

(19) “that Bob is a great racquetball player”

refers to the proposition semantically encoded in

(9) “Bob is a great racquetball player”

Supposing, as we have been, that proper names are mere labels, there is some object O such that the proposition referred to by both (18) and (19) is:

(20) That O is a great racquetball player.

In light of these points, consider the following pair of sentences:

(21) “Fred believes that Bob is a great racquetball player”,

(22) “Fred believes that Smith is a great racquetball player.”

(21) attributes a certain property to a certain proposition. The property is: *being believed by Fred* and the proposition is (20). (22) attributes a certain property to a certain proposition. As before, the property is: *being believed by Fred* and the proposition is (20).

Thus, (21) and (22) semantically encode the very same proposition. But it is obvious that, under the circumstances earlier described, what (21) *communicates* may differ in truth-value from what (22) communicates. Here two questions arise. First, exactly why (under those circumstances) do (21) and (22) differ in respect of what they communicate, given that they coincide in respect of literal meaning? Second, exactly what (under those circumstances) do those two sentences communicate?

Given what we've said so far, we can easily answer those questions, and we will now do so. We know that *semantically* "Bob" is a name. But we've already seen why, under the circumstances in question, sentences of the form "…Bob…" communicate propositions of the form: *there is a person x such that x uniquely has thus and such characteristics (viz. x is a pudgy and unathletic looking person with whom I shook hands a moment ago…) and "Bob" is x's name; moreover,…x…*

Under those circumstances, then, "Bob" is *acting* as a certain kind of existential quantifier. In other words, if we focus on what "…Bob…" is *communicating* to me, and not on what it literally means, "Bob" is acting as a certain kind of quantifier---the kind that, according to Russell's Theory of Descriptions, definite descriptions are. Let us refer to quantifiers of that kind as "Russellian quantifiers."

Consider the sentence:

(21) "Fred believes that Bob is a great racquetball player"

We've seen that, although *semantically* it is just a label, "Bob" is *functioning* (in respect of what it communicates, under the circumstances earlier described) as a quantifier.

That quantifier can be given either "wide-scope" or "narrow-scope" with respect to "Fred believes." So, speaking in terms of what (21) communicates (as opposed to what it literally mans), (21) is ambiguous between:

(23) Fred believes that: there is a person x such that x uniquely has thus and such characteristics (viz. x is a pudgy and unathletic looking person with whom I shook hands a moment ago…) and "Bob" is x's name and, moreover, x is a great racquetball player.

and

(24) There is a person x such that x uniquely has thus and such characteristics (viz. x is a pudgy and unathletic looking person with whom I shook hands a moment ago…) and “Bob” is x’s name and, moreover, Fred believes that: x is a great racquetball player.

By exactly similar reasoning,

(22) “Fred believes that Smith is a great racquetball player.”

is (in terms of what it communicates, not what it literally means) ambiguous between:

(25) Fred believes that: there is a certain person x such that x uniquely has such and such properties (viz. yesterday, from a considerable distance, I saw x leaping about in a decidedly athletic manner, and x was wearing a bright-red shirt…), and such that “Smith” is x’s name and, moreover, that x is a great racquetball player.

and

(26) There is a certain person x such that x uniquely has such and such properties (viz. yesterday, from a considerable distance, I saw x leaping about in a decidedly athletic manner, and x was wearing a bright-red shirt…), and such that “Smith” is x’s name and, moreover, Fred believes that: x is a great racquetball player.

Given that, under the circumstances in question, (21) and (22) are ambiguous in the way just described, it is no wonder that they may communicate propositions having different truth-values. Obviously (23) could be true even if (24) was false. Obviously (25) could be true even if (26) was false. This much is obvious from basic logic; no recherché results in semantics need be considered.

Given any context-–modal, moral, causal, etc.---an argument exactly similar to the one just given shows that replacing a name occurring in that context with a co-referring name cannot possibly lead from truth to falsity, at least not at the level of semantics.

III. Definite Descriptions

Extensionality says that replacing referring terms with co-referring terms never changes truth-value. Extensionality doesn't say anything about what happens when terms that don't co-refer are intersubstituted. So if, as Russell held, definite descriptions are *not* referring terms, then the apparent difference in truth-value between, for example, (5) and (6) is irrelevant to the question whether extensionality is true. But, of course, it cannot be taken for granted that definite descriptions are not referring terms. So, in order to establish extensionality, one of two contentions would have to be established, namely:

(A) Definite descriptions are not referring terms.

(B) Even if definite descriptions are in fact referring terms, intersubstituting

co-referring definite descriptions does not, contrary to initial appearances, change truth-value.

If (B) is the case, then (5) and (6) cannot differ in truth-value – even though, of course, they very much seem to. Of course, if (B) is correct, then the question whether (A) is correct becomes irrelevant to the question whether extensionality is correct.

We will thus argue for (B). And we will see that, even if (5) and (6) have the very same literal meaning, there is no difficulty accounting for the fact that they may obviously *communicate* propositions that have (or might have) different truth-values.

Here we come to a delicate issue. It is widely held that there are *two* kinds of referring terms – those that refer "directly" and those that refer "indirectly." Let me define these terms. Suppose that, addressing Fred, I say "you are tall." According to David Kaplan (1989), there is some object x such that x is identical with Fred and such that what is literally meant by my utterance is the singular proposition: *x is tall.* Appropriately generalized, Kaplan's position amounts to this. Let D be a token of an indexical expression that refers to O. (An indexical is an expression whose referent is a function of the circumstances in which it is uttered. Examples of indexicals are "I", "you", "that car", and "this bird.") In that case, in virtue of having the form "…D…", a sentence-token S has for its literal meaning a proposition that has as a constituent *O itself* – not some Fregean sense, not some concept under which O falls, and not some description that O satisfies. So if, while speaking to you, I produce a sentence-token of the form "…you…", then in virtue of that fact my utterance encodes a proposition that has you – not a concept that you satisfy, but you *per se* – as a constituent. More exactly, there is some x such that x is identical with you and such that, if I say "you are

tall", the proposition literally meant by my words is simply: *x is tall*. The concept *person to whom I am speaking* does not make it into that proposition; nor does any other concept under which you happen to fall. My utterance of "you" is thus "directly referential." Of course, the concept *person to whom I am speaking* plays an important semantic role in this context. But, if Kaplan is right, that role is only to *pick out* the entity that makes it into the relevant proposition – that concept does not itself do so.

To take another example, suppose that I point to some car x and say "that must have cost a fortune." In that case, supposing that Kaplan's position is correct, the literal meaning of my utterance is: *x must have cost a fortune.* The concept *ostended object* is no part of the literal meaning of my utterance. Obviously that concept is playing an important semantic role. But, if Kaplan is right, that role is to *pick out* a certain object. And it is that object – not the concept *ostended object* – that makes it into the proposition meant by my words.

If an expression E refers in the manner just described to some object O – in other words, if E "directly" refers to O – then E's semantic value is O itself, and not some concept that O satisfies. (By E's "semantic value" I mean what E contributes to the meaning of the sentence-tokens in which it occurs.) Under this circumstance, E refers "directly" to O in the sense that E's semantic value is O itself, and not a concept or sense that mediates between E and O.

E refers "indirectly" to O if, in virtue of having the form "…E…", a sentence (or sentence-token) S has for its meaning a proposition that has as a constituent not O itself, but rather some concept that O satisfies. So if "the inventor of bifocals" refers indirectly to Benjamin Franklin, then in virtue of having the form "…the inventor of bifocals…", a sentence encodes a proposition that contains the concept *bifocal inventor* (or *unique bifocal*

inventor) as a constituent. Many, including Frege, hold that definite descriptions are, in fact, devices of indirect reference.

In this section we will see reasons to accept following claims. The supposition that definite descriptions are directly referential is not only consistent with, but actually accounts for, the *apparent* differences in meaning between, for example, (5) and (6). That supposition also accounts for the fact that (5) and (6) *communicate* propositions that can have different truth-values. We will also see that, if definite descriptions are directly referential, then intersubstituting them never changes the truth-value of what is *literally meant* by sentences. We will wait until the next section to see what the consequences for extensionality are of supposing that definite descriptions are devices of *in*direct reference.

Let us use the term "direct referentialism" to refer to the position that definite descriptions are devices of direct reference. And let us use the term "Russellianism" to refer to the position that definite descriptions are not referring terms and are instead quantifiers (or quantifier-like expressions) of the kind that Russell described.

We must begin by noting how much direct referentialists *agree* with Russellians. Russell says that:

(27) "the king of France is bald"

means (or, at least, encodes a proposition logically equivalent with):

(28) exactly one thing x is a king of France; moreover, x is bald.

Russell thus thinks that, if (27) is to be true, then there must be exactly one king of France and any such person be bald.

Direct referentialists agree *almost* entirely with Russell. A direct referentialist will agree that if (27) is to be true, then it is necessary and sufficient that there be exactly one king of France x and that x be bald. In general, Russellians and direct referentialists both agree that "…the phi…" is true iff: something x uniquely has phi and…x…

But Russellians and direct referentialists obviously *do* disagree about the semantics of "…the phi…" Their disagreement can be understood as concerning where to put the "means that" operator in the semantic rule that gives the meaning of the phi. For Russellians, the rule that assigns meaning to "…the phi…" is this:

(29) "…the phi…"

means

(30) there is exactly one object O that has phi; moreover,…O…

For direct referentialists, the rule that assigns meaning to "…the phi…" is this:

If* there is exactly one object O that has phi, *then

(29) "…the phi…"

means

(31) ...O…

If that condition isn't satisfied, i.e. if nothing uniquely has phi, then (29) doesn't bear any proposition at all, and is thus neither true nor false.

So according to the direct referentialist,

(32) "the inventor of bifocals snored"

doesn't encode any proposition if nobody uniquely invented bifocals. But *if* exactly one person O invented bifocals, then (32) means:

(33) O snored.

So given that, in actuality, Benjamin Franklin uniquely invented bifocals, the direct referentialist says that (32) encodes the proposition:

(34) Benjamin Franklin snored.

But according to the Russellian, the proposition encoded in (32) is identical (or at least equivalent) with:

(35) exactly one object x invented bifocals; moreover, x snored.

Given that (34) doesn't entail (35), and that (35) doesn't entail (34), there thus turns out to be a significant difference between the direct referentialist's position and the Russellian's.

A consequence of the direct referentialist's position is that, for some object O (such that, as it happens, O is identical with Benjamin Franklin),

(36) “Fred believes that the inventor of bifocals snored”

encodes the proposition:

(37) Fred believes that O snored.

It follows that, for the direct referentialist,

(38) “Fred believes that the first post-master general snored”

also has (37) for its meaning.

Another consequence of the direct referentialist’s position is that, for some O (such that O is identical with the number two),

(5) “Little Timmy believes that the whole number that comes immediately after one is less than three”

has for its literal meaning the proposition:

(39) Little Timmy believes that O is one less than three.

And it follows that, for the direct referentialist,

(6) “Little Timmy believes that the number n such that n is a unique even prime if arithmetic is incomplete and is otherwise a unique positive square root of 71 is less than three.

also has (39) for its meaning.

Thus, *if* direct referentialism is correct, then (5) and (6) coincide in respect of literal meaning and, therefore, in respect of truth-value, the same being true of (36) and (38).

But, of course, there is a glaring difference in cognitive value between (36) and (38) and also between (5) and (6). There is no denying that the proposition that is *communicated* by (36) could differ in cognitive value and also truth-value from that communicated by (38) or that the proposition communicated by (5) could differ in both those respects from that communicated by (6).

For this reason, it is often assumed that the direct referentialist cannot explain why, for example, (36) and (38) differ so greatly in respect of cognitive value. This assumption is erroneous. If the direct referentialist is right, then the semantic rule for:

(32) "the inventor of bifocals snored"

is this:

(RF) *If* somebody O uniquely invented bifocals, *then* a token of "the inventor of bifocals snored" means (or is true iff): O snored. And if nobody uniquely invented bifocals, then "the inventor of bifocals snored" isn't true or false.

So if the direct referentialist is right, (RF) is thus the rule through which people who understand (32) assign meaning to it. Thus, an immediate consequence of the direct referentialist's position is that anyone who understands (32) knows that it isn't either true or false, and that it therefore isn't true, unless *exactly one object x invented bifocals and, moreover, x*

snored. Notice that the italicized proposition is identical with (35). Thus, an immediate consequence of the direct referentialist's position is that an utterance of (32) will *communicate* (35).

For the sake of argument, suppose that, if the direct referentialist were right, then the semantic rule for (32) would be:

(NR) "The inventor of bifocals snored" means (or is true iff): *Benjamin Franklin snored.*

In that case, it would indeed be inexplicable why utterances of (32) communicated (35). But (NR) obviously isn't the right semantic rule. What "the inventor of bifocals" refers to is a function of who invented bifocals. If Smith did so, then "the inventor of bifocals" refers to Smith. If Jones did so, then it refers to Jones. So if definite descriptions are devices of reference, then the semantic rule for "the inventor of bifocals" is: *if* somebody O uniquely invented bifocals, *then* "the inventor of bifocals" refers to O. Thus, if definite descriptions are devices of *direct* reference, then the semantic rule for "the inventor of bifocals" is: if somebody O uniquely invented bifocals, then "the inventor of bifocals" refers *directly* refers to O. It follows that, if definite descriptions are devices of *direct* reference, then the semantic rule for "the inventor of bifocals snored" will be (RF). We've seen why, given this last fact, it follows that, if the direct referentialist is right, a token of (32) will *communicate* (35).

Supposing that the direct referentialist's position is correct, an argument exactly similar to that just given establishes that what is communicated (not literally meant) by:

(40) "the first postmaster general snored"

will be identical or at least equivalent with:

(41) exactly one object x was a unique first postmaster general; moreover, x snored.

So, while the direct referentialist does indeed hold that, for some object O, the bare singular proposition:

(33) O snored

is the literal meaning of (40), nonetheless the direct referentialist's position is not only consistent with, but has as an immediate consequence, the fact that what is communicated by (40) will *not* be that bare singular proposition and will instead be the descriptively richer (41).

It is not *ex nihilo* that people understand the sentences that they read and hear. People must compute sentence-meaning on the basis of background information – on the basis of, in particular, a knowledge of the semantic rules that assign meanings to those sentences. Given this fact, it is an immediate consequence of the direct referentialist's position that:

(29) "…the phi…"

communicates a proposition identical, or at least equivalent, with:

(30) there is exactly one object x that has phi; moreover,…x…

In other words, it is a consequence of the direct referentialist's position that,

in terms of how it affects what is communicated (not literally meant) by sentences containing it, "the phi" functions like a Russellian quantifier. (Let "CR" be our term for the fact that the referentialist's position has the consequence just described.) It is therefore clear why the direct referentialist's position has no trouble accounting for the massive differences in cognitive value between (36) and (38), and also between (5) and (6). Given CR, it follows that, if the direct referentialist's position is right, everything that Russell's theory says about the literal meanings of sentences containing definite descriptions is true of what such sentences communicate. Russell's theory says that (36) is ambiguous between:

(36_{WS}) Somebody x uniquely invented bifocals; moreover, Fred believes that x snored

and

(36_{NS}) Fred believes that somebody x uniquely invented bifocals and, moreover, that x snored.

And Russell's theory says that (38) is ambiguous between:

(38_{WS}) Somebody x was a unique first postmaster general; moreover, Fred believes that x snored

and

(38_{NS}) Fred believes that somebody x was a unique first postmaster general and, moreover, that x snored.

Thus, a consequence of direct referentialism is that, in terms of what it communicates (not what it literally means), (36) is ambiguous between (36_{NS}) and (36_{WS}). Another consequence of direct referentialism is that (38) is ambiguous, in respect of what *it* communicates (not what it literally means), between (38_{NS}) and (38_{WS}). Neither one of (36_{NS}) or (36_{WS}) entails, or is entailed by, either one of (38_{NS}) and (38_{WS}). Thus, direct referentialism is perfectly consistent with the fact that, in terms of what they communicate, (36) and (38) may differ in truth-value. For exactly similar reasons, direct referentialism is consistent with the fact that, in terms of what *they* communicate, (5) and (6) may differ in truth-value.

Given CR, a consequence of the direct referentialist's position is that:

(42) "the inventor of bifocals is identical with the first postmaster general"

communicates a proposition identical with (or equivalent with or, at the very least, to much the same effect as) the proposition:

(43) exactly one thing x uniquely invented bifocals, and exactly one thing y was a unique first postmaster general; moreover, x=y.

So direct referentialism has no trouble accounting for the fact that, in terms of what it communicates, (42) is non-trivial. Of course, according to Russell's theory, either (43) itself, or some proposition logically equivalent with (43), is what is meant by (42).

We are seeing that direct referentialism has many of the virtues that are ascribed to the Theory of Descriptions. Direct referentialism readily explains why (5) and (6) differ in respect of cognitive value and also why (42) is non-

trivial. Further, direct referentialism readily accounts for the fact that, in terms of what utterances of it communicate, (5) is ambiguous between "wide-scope" (de re) and "narrow scope" (de dicto) readings. So if the Theory of Descriptions is to be preferred to referentialism, it cannot, contrary to what is widely assumed, be on the grounds that the latter is inconsistent with the facts concerning what is communicated by sentences containing definite descriptions---it cannot be on the grounds that the former does, whereas the latter does not, accommodate the fact that "…the phi…" communicates *there is exactly one object x that has phi; moreover,…x…*

But what is important in this context is that direct referentialism, supposing it correct, is no threat to extensionality. As we saw earlier, if direct referentialism is correct, (5) and (6) coincide in respect of literal meaning, and therefore in respect of truth-value, as do (36) and (38). It might be thought that this victory is a hollow one since, given how much they differ in respect of what they communicate, (5) and (6) must surely differ in respect of literal meaning and, consequently, that a direct-referentialist analysis of definite descriptions is a non-starter. But such a view, we have just seen, is quite misguided.

IV. The Fregean Alternative: definite descriptions as devices of indirect reference

We've considered the direct referentialist's view, according to which the sense of a definite description doesn't make it into the proposition meant by the sentence in which that definite description occurs. We've also considered the view that definite descriptions are quantifiers. And we've seen that, if we focus only these two views, then substitution-failures involving definite descriptions don't threaten extensionality. But there is a third possibility to

consider, viz. that definite descriptions are devices of indirect reference. (This was Frege's view.) Let us refer to this view as "indirect referentialism", and let us make it clear what exactly indirect referentialism comes to.

Consider the sentence:

(32) "the inventor of bifocals snored."

Here is what the indirect referentialist says about (32). First, "the inventor of bifocals" refers to Benjamin Franklin. (Thus, the indirect referentialist rejects Russell's Theory of Descriptions.) But the proposition literally meant by (32) does not have Benjamin Franklin himself as a constituent. Rather that proposition has as a constituent some concept (or "sense", as Frege put it) under which Benjamin Franklin uniquely falls. (That concept would be *inventor of bifocals* or perhaps *unique inventor of bifocals.*) More generally, according to the indirect referentialist, if a sentence S has the form "…the inventor of bifocals…", then in virtue of that fact S encodes a proposition that has as a constituent a concept C such that the referent of "the inventor of bifocals" satisfies C. (C be would be the concept *(unique) inventor of bifocals.*) Even more generally, if a sentence S has the form "…the phi…", then in virtue of that fact S encodes a proposition that has as a constituent a concept C such that the referent of "the phi" satisfies C. (C would be the concept *(unique) object having property phi.*)

If the indirect referentialist is right, it strongly *appears* to follow that extensionality is false. But I would now like to show that, so far as it is not an incoherent doctrine, indirect referentialism is not substantially different from the view that definite descriptions are quantifiers. So given that, as we have seen, Russell's theory is consistent with extensionality, the same is true of indirect referentialism, *so far as* the latter is not an incoherent, and therefore

false, doctrine.

Reference versus quantification

As we just discussed, if the indirect referentialist is right, then (32) encodes a proposition P that has the concept *inventor of bifocals* as a constituent. Supposing that the indirect referentialist is right, what must be true of that concept if (32) is to be true? That concept must be uniquely instantiated, and any instance of it must have snored. So, if the indirect referentialist is right, then the proposition meant by (32) is true just in case

(35) exactly one object x invented bifocals; moreover, x snored.

But (35) is the proposition that, according to the Russellian, is meant by (32). Of course, the essence of the Russellian's position is definite descriptions are *not* referring terms. So given only what we've seen thus far, it would appear that indirect "referentialism" is really the view that definite descriptions are quantifiers.

There are additional reasons to believe that indirect referentialism (so-called) collapses into the view that definite descriptions are *not* referring terms. Consider the sentence:

(44) "Benjamin Franklin snored."

Let P be the proposition meant by (44). Given Kripke's (1980) arguments, we know that there is some object O (such that O is Benjamin Franklin) such that, given any possible world w, P is true in w just in case O snored; in other words, P is true just in case (33) is true. (This, of course, is consistent with

the supposition that P and (33) are identical.) More plainly, the proposition meant by (44) is true in any possible world just in case Benjamin Franklin snored in that world. This point is easily generalized. There is some object O (such that O is Benjamin Franklin) such that, given any sentence S having the form:

(45) "…Benjamin Franklin …"

the proposition meant by S is true in a world W just in case…O…(i.e. just in case O has property…x…).

Of course, we know that "Benjamin Franklin" is a paradigm-case of an expression that refers to Benjamin Franklin. (It is also a paradigm case of an expression that *directly* refers to Benjamin Franklin.)

Now let us consider the expression "the inventor of bifocals." If the *in*direct referentialist is right, then the proposition meant by (32) is true in many worlds where (33) is false, and it is false in many worlds where (33) is true.

These points are easily generalized. Let S be any sentence of the form:

(29) "…the phi…",

Let P_{IR} be the proposition that, *if the indirect referentialist is right*, is literally meant by S. Let P_{TD} be the proposition that, if the Theory of Descriptions is right, is meant by S. Finally, let P_{DD} be the proposition that, if the direct referentialist is right, is meant by S.

There are many worlds where P_{DD} is true and where P_{IR} is false, and many worlds where P_{DD} is false and where P_{IR} is true. At the same time, P_{IR} is true in exactly the same worlds as P_{TD}. Of course, this last point does

not by itself show that P_{IR} is *identical* with P_{TD}. But it does strongly suggest, if it doesn't outright entail, that P_{IR} is identical with *some* quantified generalization of the kind that, if Russell's theory is correct, is meant by (32).

Thus, given what we've seen thus far, there are many semantic similarities between indirect reference (so-called) and quantification and, in fact, there don't seem to be any significant semantic differences between them. Further, there are clear semantic differences between, on the one hand, expressions that are *clearly* devices of reference (e.g. "Benjamin Franklin") and, on the other hand, expressions that refer "indirectly."

It is also worth pointing out that, from a proof-theoretic standpoint, P_{IR} is not significantly different from P_{TD}: each entails, and is entailed by, exactly the same propositions as the other.

In fact, not only are P_{TD} and P_{IR} equivalent; they seem to coincide in respect of their more fine-grained inferential properties. Permit me to say what I mean by this. Because they are true in exactly the same possible worlds (namely, all of them), *1+1=2* and *triangles have three sides* are equivalent. But the *way* in which *1+1=2* entails *for some number n, 1+n=2* is very different from the way in which *triangles have three sides* entails that same proposition. (If *1+1=2* is one's starting point, a single application of the rule of existential generalization is enough to prove *for some number n, 1+n=2*. But this is not the case if *triangles have three sides* is one's starting point.) So even though they coincide in respect of *what* they entail, *1+1=2* and *triangles have three sides* do not coincide in respect of *how* they entail it, and they therefore do not coincide in respect of what we might describe as their "fine-grained" inferential properties. So, even though they are equivalent, they are not proof-theoretically interchangeable.

By contrast, P_{TD} and P_{IR} *do* seem to be thus interchangeable. Let P be the proposition *at least one person invented bifocals*. Not only do P_{TD} and

P_{IR} both entail P – they do so in similar, if not identical, ways.

Permit me to clarify this last point. Suppose that you ask an indirect referentialist what the semantic rule for (32) is. If he answers by saying:

(*) the proposition affirmed by the sentence “the inventor of bifocals snored” is true just in case the inventor of bifocals snored,

his answer is ambiguous, as it could mean either:

(**) the proposition affirmed by the sentence “the inventor of bifocals snored” is true just in case the person who *in fact* invented bifocals snored; in other words, if x is a person who uniquely invented bifocals, then the proposition affirmed by “the inventor of bifocals snored” is true just in case x snored,

or (*) could mean:

(***) the proposition affirmed by the sentence “the inventor of bifocals snored” is true just in case somebody x uniquely invented bifocals and, moreover, x snored.

Until the indirect referentialist identifies the intended disambiguation, he hasn’t told what his semantics is. Of course, (***) is the intended disambiguation – after all, (**) is what the *direct* referentialist believes the relevant semantic rule to be. But supposing that (***) gives the semantics of (32), the proposition meant by that sentence coincides precisely in respect of its proof-theoretic properties with the proposition that, according to the Theory of Descriptions, is meant by that sentence. In fact, (***) is itself

identical with the rule, that according to that theory, gives the semantics of (32). So it is by no means easy to see any proof-theoretic differences between P_{TD} and P_{IR}; and, in light of that fact, it is not easy to see why, if the indirect referentialist is right, P_{IR} is anything other than a quantified generalization.

McCawley writes (1971: 23): "quantifiers [are] elements of logical structure that correspond to words like *all*, *some*, and *most*, which say which or how many of some set of things have a given property." So "no inventor of bifocals" is a quantifier because the sentence "no inventor of bifocals is tall" answers the question "how many inventors of bifocals are tall?", and "some inventor of bifocals" is a quantifier for the same reason (*mutatis mutandis*). In general, a quantifier is an expression that, when embedded in a sentence, answers a question of the form "how many phi's are psi's? "If the indirect referentialist is right, (32) is logically equivalent with the proposition: exactly one thing invented bifocals, and any such thing is tall. So, if the indirect referentialist is right, then (32) answers *two* questions of the form "how many phi's are psi's? " It answers the question: "how many things are bifocal-inventors?" and it also answers the question "how many bifocal-inventors snored?" So if the indirect referentialist is right, "the inventor of bifocals" would seem, at the very least, to be quantifier-like in some important respects, the same being true of every other definite description.

What we are seeing is that, if the indirect referentialist is right, the semantics of "the inventor of bifocals" is, at the very least, extremely similar to the semantics that the Russellian ascribes to it. We are also seeing that, if the indirect referentialist is right, "the inventor of bifocals" is semantically very different from any instance of an expression that *clearly* refers to Benjamin Franklin (e.g. "Benjamin Franklin" or a token of "that guy over there" that is accompanied by an ostension of Benjamin Franklin).

Of course, it would be an overstatement to say that clear-cut cases of reference have *no* significant similarities with cases of indirect reference. But to the extent that there are such similarities, those same similarities are also found between clear-cut cases of reference, on the one hand, and clear-cut cases of quantification, on the other. For that reason, those similarities give no more weight to the position that definite descriptions are devices of reference than they do to the position that they are quantifiers.

Permit me to substantiate this. Given the open sentence "x snored", a complete sentence can be formed by replacing the variable with either "Benjamin Franklin" or "the inventor of bifocals"; and this is *consistent with* the idea that "Benjamin Franklin" and "the inventor of bifocals" are semantically similar. (If the variable were replaced with the word "and", the result would be nonsense, and this is plausibly seen as having some basis in the semantics of the word "and.") But this is not enough to show that (so called) indirect reference really is reference. After all, a meaningful sentence results if the variable is replaced with a quantifier (e.g. "some person", "no frog").

Taken as a complete utterance, the expression "and snores" is meaningless. The reason is that "and" doesn't combine with "snores" in the right way: it has the wrong combinatorial properties---the wrong syntax, in other words. It doesn't have the combinatorial-syntactic properties had by, for example, "Benjamin Franklin." And the reason that both "the inventor of bifocals snored" and "Benjamin Franklin snored" are meaningful is that "the inventor of bifocals" and "Benjamin Franklin" *do* have similar combinatorial-syntactic properties – the reason is not that they have the similar *semantic* properties. After all, "some man", no less than "Benjamin Franklin", has the right syntactic properties ("some man snored" is no less meaningful than "Benjamin Franklin snored"). And if Russell and Frege have taught us

anything, it is that “some man” is semantically very different from any singular term. Of course, “some man” is, in respect of its *syntactic* properties, similar to “Benjamin Franklin.” But reference is a semantic, not a syntactic, relation: when you say that, for some x, “Benjamin Franklin” *refers* to x, you are making a statement about the semantics of “Benjamin Franklin”, and not (at least not primarily) about its syntax.

For the sake of argument, let us suppose in this paragraph that the indirect referentialist is right. Given what we said in the preceding two paragraphs, it is clear that, while there are indeed similarities between “Benjamin Franklin” and “the inventor of bifocals”, comparable similarities are found to hold between “Benjamin Franklin” and various quantifiers (e.g. “some man”, “no person”). For that reason, the existence of those similarities no more supports the claim that “the inventor of bifocals” is a referring term than it supports the claim that it is a quantifier. Moreover, the similarities between “Benjamin Franklin” and “the inventor of bifocals” seem to be syntactic, not semantic, in nature. But given that reference is a semantic function, those similarities don’t support the contention that “the inventor of bifocals” (supposing, as we momentarily are, that it is a device of indirect reference) is anything other than a quantifier.

V. Why indirect referentialism is committed to an incoherent semantic view

There is yet another reason to think that indirect reference (so-called) is identical with Russellian quantification – that, more precisely, indirect referentialism is merely a confused version of Russellianism.

Remember that the essence of the indirect referentialist’s position is as follows (it is to be supposed that O is the one and only object having property phi):

If S is a sentence of the form,

**(29) "…the phi…", **

then, in virtue of that fact, S has as a constituent, not O itself, but rather some concept C that O uniquely satisfies, that concept being identical with (or, at least, along the lines of): *unique object having phi*.

We know from Frege that, taken in conjunction with an extremely plausible semantic principle, indirect referentialism has some extremely counterintuitive consequences. The principle in question is this: the meaning of a complex expression is a function of the meanings of its constituents. This is known as the "principle of compositionality" (or simply "compositionality"). A related principle is that the referent of a complex referring term is a function of the referents of its constituents. We will refer to this as "special compositionality." Obviously compositionality and special compositionality are extremely plausible principles.

The indirect referentialist (unlike the Russellian) holds that "the inventor of bifocals" and "the first post-master general" are referring terms and that, given certain facts about history, they therefore co-refer. As we've noted, if indirect referentialism is right, then:

(36) "Fred believes <u>that the inventor of bifocals snored</u>"

and

(38) "Fred believes <u>that the first postmaster general snored</u>"

differ in meaning and also (potentially) in truth-value. And supposing that the indirect referentialist is right about this, the reason is that the underlined part of (36) refers to a different proposition from the underlined part of (38).

Given only these points, the indirect referentialist's position seems plausible. But, as Frege noted, an implausible consequence soon follows. (Because Frege was himself an indirect referentialist, he accepted this consequence.) For the sake of argument, suppose that, as the indirect referentialist holds, the underlined part of (36) does not refer to the same proposition as the underlined part of (38). Given special compositionality, it follows that the occurrence in (36) of "the inventor of bifocals" doesn't co-refer with the occurrence in (38) of "the first postmaster general."

The reason that, if the indirect referentialist is right, the underlined parts of (36) and (38) do not co-refer is obviously that the concept associated "the inventor of bifocals" is different from the concept associated with "the first postmaster general." Given special compositionality, it follows that the occurrence in (36) of "the inventor of bifocals" refers to the *concept* with which it is associated (viz. the concept *(unique) inventor of bifocals*). For exactly similar reasons it follows that, if the indirect referentialist is right, the occurrence in (38) of "the first postmaster general" refers to the concept *(unique) first postmaster general*.

Thus, if the indirect referentialist is right, the occurrence in (36) of "the inventor of bifocals" doesn't co-refer with the occurrence of that same expression in (32): the latter occurrence refers to Benjamin Franklin; the former refers to the concept *(unique) bifocal inventor*. Given obvious extensions of this reasoning, it follows that, if the indirect referentialist is right, the occurrence of "the inventor of bifocals" in:

(46) “it is true that the inventor of bifocals snored”

does not refer to Benjamin Franklin and that it *does* refer to the concept *(unique) bifocal inventor*. It also follows that, if the indirect referentialist is right, the occurrence of that same expression in:

(47) “it is true that it is true that the inventor of bifocals snored”

refers not to Benjamin Franklin, nor to the concept *(unique) bifocal inventor*, but rather to a concept *of* that concept.

Of course, it is intuitively very hard to believe that, in (36), the inventor of bifocals doesn’t refer to the inventor of bifocals. And it is very hard to believe that, in (47), “the inventor of bifocals” doesn’t co-refer with its homonym in *either* (36) or (32).

But I believe that, in addition to being implausible, it is incoherent to suppose that definite descriptions undergo reference-shifts of the kind just described. The reason for this is can be understood in terms of the Theory of Descriptions. As we’ve discussed, the heart of that theory is that:

(TD) if S is a sentence of the form “…the phi…”, then in virtue of that fact S has for its meaning either the proposition *something O uniquely has phi and…O…*or some other, logically equivalent, quantified generalization.

One virtue of the Theory of Descriptions is that there is some *one* rule (namely TD) such that, given *any* occurrence of a given definite description, the semantics of that occurrence is given by that rule. So, for example, Russell’s theory has the virtue that there is some *one* rule such that the

semantics of the occurrence of "the inventor of bifocals" in each of (32), (36), (46), and (47) is given by that rule.

Before we proceed, there is a nuance that we must address. Let S be a sentence that contains an occurrence of "the phi" and suppose that, in S, "the phi" occurs within the scope an operator that forms molecular sentences out of atomic ones. In that case, (TD) assigns (at least) two different meanings to S and Russell's theory thus says that S is ambiguous between those two meanings. For example, as we've discussed, if Russell's theory is right, then (36) is ambiguous between (36_{WS}) and (36_{NS}). But this doesn't mean that, if Russell's theory is right, there is not some *one* rule in terms of which the semantics of occurrences of definite descriptions can be understood. On the contrary, it means that there *is* such a rule---albeit one that sometimes assigns multiple meanings to a given sentence.

In light of these points, let us turn our attention back to indirect referentialism. We've seen that, if that doctrine is right, then "the inventor of bifocals" refers to one thing in (32), a second thing in (36), a third thing in (47), and so on.

Let us now take the next, and final, step in our argument. If indirect referentialism is right, then given any occurrence O of "the inventor of bifocals" in any sentence S, there is some *single* rule saying how S's truth-conditions are affected by virtue of its containing O. Permit me to clarify this. As we've noted, if indirect referentialism is right, then in virtue of having the form "…the phi…" a sentence S has the concept *phi* as a constituent and, further, S is true just in case something O uniquely instantiates *phi* and…O… So, for example, if indirect referentialism is right, then (32) is true just in case something O uniquely instantiates the concept *bifocal inventor* and O snores. Thus, the heart of the indirect referentialist's position is given by the rule:

(IR) If a sentence S has the form "…the phi…", then in virtue of that fact S has for its literal meaning a proposition that is true just in case something O uniquely has phi and, moreover,…O…

Of course, if (IR) is a correct characterization of the semantics of definite descriptions, then some sentences containing definite descriptions are, in virtue of that fact, ambiguous. For example, (36) would be ambiguous between two propositions – one of them logically equivalent with (36_{WS}), the other logically equivalent with (36_{NS}). But (to echo what we said a moment ago) this doesn't mean that, if indirect referentialism is correct, there is not some *one* rule in terms of which the semantics of any occurrence of a definite descriptions is to be understood. On the contrary, it means that there *is* such a rule (namely, IR) – albeit one that sometimes assigns multiple meanings to a given sentence.

So if indirect referentialism is right, there is some *one* rule that correctly characterizes the semantics of the occurrence of "the inventor of bifocals" in (32), and that also correctly characterizes the semantics of the corresponding occurrences in (36), (46), and (47). A consequence is that "the inventor of bifocals" is, from a semantic viewpoint, doing the same thing in all of those sentences. In general, supposing that indirect referentialism is correct, then given any occurrence O of any definite description in any sentence S, O's effect on the truth-conditions of S are to be understood in terms of IR: no matter where "the phi" occurs in a sentence – no matter how great the number of operators within whose scope that occurrence falls---IR tells you how that sentence's truth-conditions are affected by its containing that occurrence. A consequence is that (supposing that indirect referentialism is correct) any two occurrences of a given definite description are, from a semantic point of

view, doing precisely the same thing, even if one of them falls within the scope of a greater number of operators than the other.

Now we can close the argument. We have seen (and Frege himself made it clear) that, if indirect referentialism is right, definite descriptions are ambiguous – that, for example, "the inventor of bifocals" refers to one thing in (32), a different thing in (36), a third thing in (47), and so on. But we have just seen that, if indirect referentialism is right, then "the inventor of bifocals" is, semantically speaking, doing the same thing in every one of those sentences. That expression has the same effect on the truth-conditions of (32) that it has on (36) and on (47) and also on:

(48) "it is true that Fred believes that Mary doubts whether the inventor of bifocals snored."

Thus, if indirect referentialism is right, then *a given expression can, from a semantic viewpoint, do the exact same thing in two different contexts while referring to different objects in those contexts*. This has the consequence that the semantic facts don't fix the referential facts: two expression-occurrences can be semantically identical but referentially distinct. But, given that reference is a semantic function *par excellence*, that amounts to the absurd view that the semantics of an expression can change without changing.

VI. Reference versus quantification (revisited)

I would now like to develop the argument outlined in the previous section. There is some object O (such that O is identical with Benjamin Franklin) such that "Benjamin Franklin" refers to O. What does this mean? Part of what it means is that (given what the remaining semantic rules of

English are) if you wish to attribute property phi to O, one viable way of doing so is to say: “Benjamin Franklin has phi.” Surely it would be absurd to say that “Benjamin Franklin” referred to O but that “Benjamin Franklin has phi” didn’t attribute phi to O. So for “Benjamin Franklin” to refer to O, it is surely necessary that sentences of the form “Benjamin Franklin has phi” attribute phi to O. It is also sufficient. Suppose (what is in fact true) that, in virtue of having the form, “Benjamin Franklin has phi” a sentence attributed phi to O. It is hard to see how, under that circumstance, it could be denied that, at least in the context of a sentence having the form just described, “Benjamin Franklin” referred to O.

Of course, these points don’t constitute an exhaustive analysis of the concept of reference. But they give some weight to the idea (implicit in the work of Frege 1874 and Tarski 1972, and explicitly stated by Dummett 1965, Tughendat 1962, and others) that for E to refer to O is for it to be the case that occurrences of E have certain effects on the truth-conditions (or meanings) of sentences comprising those occurrences. (More precisely, for E to refer to O is for it to be the case that, in virtue of having the form “…E…”, a sentence says that:…O…In other words, it is for it to be the case that, in virtue of having that form, a sentence attributes to O the property of being a thing x such that…x...) If this plausible idea is at all correct, then whether an occurrence of an expression refers to some object is fixed, at least in part, by how that occurrence affects the truth-conditions (or meanings) of the sentences which comprise it. We’ve seen that, if indirect referentialism is correct, then two expression-occurrences can have the very same effect on the truth-conditions of their host-sentences *and yet refer to completely different things.* Thus, if indirect referentialism is correct, what an expression refers to is independent of what it is doing semantically, and reference therefore isn’t fixed by semantics. But, to echo what we said earlier, such a position borders

on absurdity, given that nothing is more distinctively semantic than reference.

Notice how similar (IR) is to (TD). This is consistent with our point that indirect referentialism is really just a form of Russell's Theory of Descriptions. But Russell's theory has a distinct advantage over indirect referentialism. Russell's theory denies that "the inventor of bifocals" is a referring term. So Russell's theory is not required by special compositionality to hold that "the inventor of bifocals" refers to one thing in (32), a second thing in (36), and so on. (After all, special compositionality concerns referring terms, not non-referring terms such as quantifiers.) And because it denies that definite descriptions refer, Russell's theory doesn't have the implausible consequence that two expressions that are semantically doing the same thing can nonetheless have different referents. By contrast, indirect referentialism implausibly says that each definite description is ambiguous (infinitely ambiguous, in fact, given that there is no limit to the number of operators in whose scope an occurrence of a definite description can fall). And indirect referentialism also demands (incoherently, we have argued) that two expression-occurrences can be identical in their semantic properties while differing in their referential properties.

We've seen that, if Russell's theory is correct, definite descriptions are not referring terms and, consequently, the (putative) difference in truth-value between (e.g.) (5) and (6) is consistent with extensionality. We've also seen that indirect referentialism is not substantially different from Russell's Theory of Descriptions and that, to the limited extent that those doctrines *are* different, indirect referentialism is an implausible and arguably incoherent one. Thus, even though it is true that, if indirect referentialism were correct, intersubstitutions of co-referring terms would change meaning and truth-value, that fact does not threaten extensionality.

Indirect referentialism a conflation of several distinct notions

The idea that definite descriptions are devices of indirect reference is an incoherent synthesis of a number of otherwise estimable views. The semantics of "the king of France is bald" is obviously such that an utterance of that sentence is true just in case, at the time of utterance, *there is exactly one king of France x and, moreover, x be bald*. So anyone who knows the semantics of that sentence, i.e. anyone who understands it, knows that an utterance of it is true just in case the same holds of the italicized proposition, a consequence being that such an utterance *communicates* that proposition. In general, an utterance of "…the phi…" *communicates* a proposition that is true just in case *the concept (or property) phi is uniquely instantiated by some object O and…O…*And this fact about what such utterances communicate obviously has a semantic basis. Given only these facts, it seems reasonable, if not positively *de rigueur*, to take the view that what is *literally meant* by such an utterance is equivalent with an existence-claim of the kind just described. And given that view, it readily follows that, in virtue of having the form of "…the phi…", a sentence has for its literal meaning a proposition having as a constituent the property phi. It follows, for example, that (32) has for its literal meaning a proposition having as a constituent, not Benjamin Franklin himself, but rather some property (or concept) that he uniquely instantiates.

Then there is the fact that definite descriptions *seem* very much to be devices of reference. It is hard to believe that "the number that is identical with two" (or simply "the number two") doesn't refer to the number two.

The facts described in the second to last paragraph seem to rule out direct referentialism. The fact described in the paragraph just before this one rules out Russell's Theory of Descriptions. Out of the three possibilities as to what

definite descriptions could be – devices of direct reference, devices of *in*direct reference, and quantifiers – only the second seems to be on the right side of the points made in the last two paragraphs. Indirect referentialism thus seems to be a highly reasonable semantic position.

But we saw that, to the extent that it isn't incoherent, indirect referentialism collapses into Russell's theory. This shows that there must be a non-sequitur in the argument just presented. Given what we said earlier, that non-sequitur is not to hard to identify. Let S be an arbitrary sentence of the form "…the phi…". While it is true that S communicates a proposition of the kind earlier described, and while it is true that this fact has a basis in the semantics of that sentence, it doesn't follow such a proposition is what is literally meant by S. For argument's sake, suppose that (as Strawson (1950) held), the semantic rule for S is this: *If something O uniquely has phi, then what is literally meant by S is*:…*O*…In that case, as we saw, S will *communicate* a proposition equivalent with the one that the indirect referentialist believes to be S's literal meaning, even though what is *actually* meant by S will not be that proposition or even one equivalent therewith. Once it is taken into account that, to understand a sentence, one must work through the semantic rules that assign it meaning, i.e. once a sentence's *pre*-semantics is taken into account, it becomes possible to accommodate the pretheoretic intuition that definite descriptions are referring terms (that, for example, "the number two" refers to two), while also accommodating the relevant facts about what is communicated by sentences of the form "…the phi…" Once pre-semantics is given its due, indirect referentialism proves not only *not* to be *de rigueur* but also to embody a serious confusion---a failure to distinguish between semantics and pre-semantics---and thus to be a decidedly unattractive semantic position.

VII. Quantifying In

The following sentence certainly seems to be meaningful:

(49) “John believes that Hesperus is lovely.”

And (49) certainly seems to entail:

(50) “There is some x such that John believes that x is lovely.”

But Quine (1953) says that, even if (49) is true, (50) is *not* true and, indeed, that (50) is meaningless and is thus neither true nor false. Quine says that, in general, sentences which involve “quantifying into” epistemic contexts are meaningless. Why does he think this?

Suppose for argument’s sake that (50) *is* meaningful. In that case, the open sentence “John believes that x is lovely” is true for some values of x and false for others. We can produce a closed sentence of that form by replacing the variable with a referring term. Presumably, such a closed sentence is true just in case the expression that we substitute for the variable denotes an object having the attribute expressed by “John believes that x is lovely.” Obviously “Hesperus” denotes such an object. Since “Hesperus” denotes such an object, so does “Phosphorous”, given that those expressions co-refer. Therefore, if (49) is true, then the following must also be true:

(51) “John believes that Phosphorous is lovely.”

So *if* (50) is meaningful, then (49) entails (51). But, says Quine, (49) does not entail (51). After all, Quine plausibly says, (49) and (51) might differ in

truth-value. Thus, if we grant that (50) is meaningful, then a true sentence, namely (49), entails a false one, namely (51). We must therefore conclude that (50) is *not* meaningful and that, in general, quantifying into epistemic contexts results in meaninglessness.

But, contrary to what Quine says, (50) very much *does* seem to be meaningful, and it is not hard to identify the fallacy in Quine's argument. We have seen that, contrary to what Quine assumes, (49) *does* entail (51). After all, there is some x (such that x is identical with Venus and is therefore identical with each of Hesperus and Phosphorous) such that both sentences encode the proposition: *x is lovely*. We have also seen why, that fact notwithstanding, (49) and (51) may *communicate* propositions that differ in truth-value.

The Smullyan-Quine debate

Quine believes that modal concepts (e.g. the concept of necessity) are incoherent and that the use of modal connectives (e.g. "necessarily") therefore has no place in coherent discourse. Consider the sentence:

(52) "necessarily, 9 is larger than 7."

There is little doubt that, if it is meaningful, (52) is true. Supposing that (52) is in fact true, that is because 9 has the property expressed by the open-sentence:

(53) "necessarily, x is larger than 7. "

Presumably, a true sentence results if we replace the variable with any expression (of English) that co-refers with “9.” But this is not the case, given that

(54) “necessarily, the number of planets is greater than 7”

is a false sentence that results from just such a substitution.

This argument, it is clear, coincides with, and therefore involves the same fallacy as, Quine’s argument concerning quantification into opaque contexts.

Like us, Raymond Smullyan holds that Quine’s argument is fallacious and that modal statements are meaningful. But Smullyan’s argument does not coincide with ours. A believer in Russell’s Theory of Descriptions, Smullyan holds that definite descriptions are quantifiers and that (54) is therefore not what results when a referring term in (52) is replaced with a co-referring term. Given his acceptance of Russell’s position, Smullyan is bound to say, and does say, that (54) is ambiguous between wide-scope and narrow-scope readings, these being, respectively:

(54_{WS}) There is some number x such that there are x planets, and necessarily x is greater than 7,

and

(54_{NS}) Necessarily, there is some number x such that there are x planets and x is greater than 7.

Of course, (54_{WS}) is true and (54_{NS}) is false, and it thus ceases to be a

mystery why (on at least one reading) (54) differs in truth-value from (52).

M.Wreen (1974) rejects Quine's counterintuitive view that modal expressions are categorically meaningless. Because Wreen also rejects the Theory of Descriptions, he rejects Smullyan's analysis, and he proposes the following alternative to the latter. We must distinguish between two kinds of reference, direct and indirect. (Wreen uses the term "attributive", instead of "indirect.") If the definite description in (54) is read referentially, then (54) means simply:

(54_{DR}) necessarily, 9 is greater than 7.

If the definite description in (54) is read attributively, then (54) means (or, in any case, is logically equivalent with):

(54_{AR}) necessarily, whatever the number of planets happens to be, that number is greater than 7.

(54_{DR}) is true and (54_{AR}) is false, showing why, on at least one of its readings, (54) differs in truth-value from (52).

Can Wreen's analysis be accepted? We have seen that indirect (so-called) reference is really quantification, so far as the concept of indirect reference is a coherent one. Since, according to Wreen's analysis, definite descriptions are *not* quantifiers and *are* devices of indirect *bona fide* reference, it follows that Wreen's analysis is incoherent. It also follows that, so far as Wreen's analysis *is* coherent, it isn't meaningfully different from Smullyan's, contrary to what Wreen himself says.

There is another problem with Wreen's analysis. If Wreen is right, definite descriptions are *semantically* ambiguous between (direct) referential

and attributive meanings. But in light of a persuasive argument given by Kripke (1977b), this is probably not the case. The word “dumb” is clearly ambiguous. For this reason, says Kripke, we’d expect there to be languages L such that the L-translation of “dumb” (incapable of speaking) differed from the L-translation of “dumb” (unintelligent). Kripke is right – there are many such languages. In general, translations do not, at least not categorically, preserve semantic ambiguities: translations tend to disambiguate semantically ambiguous expressions. So if an expression of one language is semantically ambiguous, it is probably not the case that, in each other language, that expression is translated by some one, comparably ambiguous expression. Thus, if (54) were *semantically* ambiguous in the way required by Wreen’s analysis, we would expect there to be a language L such that the L-translation of (54), read *de re*, was different from the L-translation of (54), read *de dicto.* But, Kripke concludes, he seriously doubts whether there is such a language. Again, Kripke is right: as a matter of empirical fact, there is no language satisfying the condition in question. Given any language L, the L-translation of (54) is ambiguous in exactly the same way as (54), suggesting that (54) is not *semantically* ambiguous.

Most importantly, given the points made in section V, there is no *need* to suppose that definite descriptions are ambiguous in the way that Wreen proposes---no need to suppose that (54) is semantically ambiguous between *de re* and *de dicto* readings. We may suppose that “the number two” refers to the number two and that, in general, definite descriptions are exactly what they appear to be: singular terms. Given this, the facts about cognitive significance – about what is communicated by (54) and by other sentences containing definite descriptions – can be explained entirely in terms of the processes in which we must engage to *compute* literal meaning.

VII. Kripke's Pierre-paradox

Kripke (1970) identified an important puzzle through the following clever thought-experiment. Pierre grows up in Paris. He speaks only French and not a word of any other language. He's seen French movies, and read French books, about London. On the basis of this information, plus his knowledge of the French language, he rationally comes to assent to the French sentence:

(55) "Londres est jolie."

Of course, Pierre understands (i); he knows what it means. And given the information at his disposal, he is quite right to assent to it.

The English translation of (55) is

(56) "London is pretty."

Of course, both sentences encode the proposition:

(57) London is pretty.

One day, Pierre is kidnapped. His kidnappers keep him hostage in a rundown, ugly part of London. Pierre learns English. But he learns it *not* by having English sentences translated into his native French, and instead learns it as one learns a first language.

On the basis of the information available to him, plus his newly acquired knowledge of English, Pierre comes to assent to the English sentence:

(58) "it is not the case London is pretty."

Pierre understands (58). And given the information at his disposal, he rationally assents to it.

Here is the problem. Pierre rationally assents to (55). He also rationally assents to (58). But the proposition semantically encoded in (55) is the negation of the proposition encoded in (58). So Pierre assents *both* to a proposition *and* its negation *and* he is not in the slightest bit irrational.

But how is this possible? After all, to be irrational just *is* to countenance some inconsistent set of propositions. In any case, this is the age old presumption.

Further, we cannot use the sense/reference distinction to solve the problem. Being translations of each other, “London” and “Londres” cannot differ in sense. So even if Frege’s semantics is correct, it won’t solve the problem. And being translations of each other, “London” and “Londres” cannot differ in quantificational content. So even if Russell’s semantics is correct, this won’t solve the problem. Kripke has thus shown that, even if Frege’s views and Russell’s views concerning the semantics of proper names are correct, this does not settle the problems those views are meant to settle.

More importantly in the present context, Kripke has also identified an important puzzle: Pierre *rationally* assents to sentences that he correctly understands and that negate each other. I would now like to show that some of our earlier remarks might give us some insight into how to solve this puzzle.

The semantic value of each of “London” and “Londres” is London. In other words, each expression refers to directly to London, and neither therefore has any semantic content not had by the other. But, for reasons that we’ve already considered in connection with (7) and (9), the meaning of “London” must be *given* to Pierre through some proposition that is descriptive. Under the circumstances described, that proposition is along the

lines of: there is some awful place x, such that I currently imprisoned in x and such that, in x, there is rotting garbage in the streets; moreover, “London” is the English label for x.

So even though, at the level of semantics, “London” means only *London*, and has no descriptive content (or, any rate, no descriptive content not equally had by “Londres”), nonetheless what “London is pretty” communicates to *Pierre* is along the lines of:

(59) there is an awful place x such that I am currently imprisoned in x, and such that there is rotting garbage in x’s streets, and such that the occupants of x are surly; moreover, x is pretty.

By exactly similar reasoning, what “Londres est jolie” communicates to Pierre (not what it semantically encodes) is along the lines of:

(60) there is a wonderful place x such that that I’ve read about x on thus and such occasions, and such that, in x, there are Beefeater guards as well as a big clock-tower; moreover, x is pretty.

To close the argument, we must deal once again with a subtlety relating to scope. Suppose Pierre hears somebody say “London is pretty.” Pierre violently recoils: “No! It is *not* the case that London is pretty!” We’ve agreed that, under the circumstances described, (59) is the proposition that “London is pretty” communicates to Pierre. But this leaves it open what “it is not the case that London is pretty” communicates to him. For we can give the negation-operator either wide or narrow-scope. The wide scope reading is:

(61) it is not the case that: there is an awful place x such that I am currently

imprisoned in x, and such that there is rotting garbage in x's streets, and such that the occupants of x are surly; moreover, x is pretty.

The narrow-scope reading is:

(62) there is an awful place x such that I am currently imprisoned in x, and such that there is rotting garbage in x's streets, and such that the occupants of x are surly; moreover, it is not the case that x is pretty.
.

It is clear that, under the circumstances, (62) is more likely than (61) to be the proposition that Pierre associates with "it is not the case that London is pretty." After all, Pierre doesn't deny that there is an awful place x such that the people in x are surly (and so on) – what Pierre denies is that x is pretty.

But here is the main point. The conjunction of (60) and (62) is not a self-contradiction, and neither, for that matter, is the conjunction of (60) and (61). The proposition that Pierre associates with "Londres est Jolie" is (60), and the proposition that he associates with "it is not the case that London is pretty" is either (61) or (62). Given what the propositions are that, under the circumstances in question, Pierre associates with "Londres est jolie" and "London is not pretty", he is obviously in no way guilty of self-contradiction in virtue of accepting both of them. We don't need to posit "propositional guises" or other dubious entities to understand what is going on. We don't deny that "London is pretty" and "Londres est jolie" have the same literal meaning. And we don't reject the age-old view that assenting to inconsistent propositions constitutes irrationality. After all, (62) is *not* inconsistent with (60), and neither is (61).

At first, Kripke's story seems to make it *de rigueur* to jettison, or at least

qualify, the deeply entrenched presumption that, given an arbitrary object x and an arbitrary property phi, the proposition *x has phi and x does not have phi,* is analytically false and therefore such that it cannot be rationally accepted. Fortunately, it isn't necessary to qualify that presumption, let alone jettison it. All that is necessary, we have seen, is to give due weight to the fact that the information through which Pierre assigns meaning to the sentence "London is pretty" differs from the information through which he assigns meaning to the sentence "Londres est Jolie."

According to Nathan Salmon, however, Kripke's parable *does* show that, under at least some circumstances, it is rational to accept the proposition *x has phi and x does not have phi*. Given that phi may be the property of not being identical with x, a consequence of Salmon's view is that the information at a given person's disposal may give that person as much reason to accept the proposition that $x\frac{4}{4}x$ as it does to accept the proposition that x=x. Salmon duly accepts this consequence. Given the presumption that *x=x* has an epistemic probability of 100%, a consequence of this consequence is that there *do* exist objects that are not self-identical. True to his own premises, Salmon duly holds that, indeed, there do exist objects that aren't self-identical. Supposing Salmon to be right about this, and supposing the principles of logic to be correspondingly revised, it's hard to see on what grounds one could deny the existence of square circles, or of even primes greater than two, or of anything else that (relative to pre-Salmonian logic) qualifies as a "surd", to use the medieval term. Once again, Salmon is true to his own premises, and duly holds that there *do* exist square circles, even primes, greater than two, etc.

Overlooking the existence of the *pre*-semantic implicatures that we have been examining, Salmon assumes that, setting aside the *post*-semantic implicatures studied by Grice, what sentence-tokens literally mean is in

lockstep with what they convey. ("Presemantic implicature"$=_{DF}$ a deviation between literal and communicated meaning that results from the influence of the information on the basis of which one figures out what is literally encoded in the expression-tokens that one encounters. "Post-semantic implicature"$=_{DF}$ a deviation between literal and communicated meaning that results from the interaction of an *already* ascertained literal meaning with facts about the context.) These being his starting assumptions, Salmon is indeed left with no choice but to take the deeply revisionist view that $x \neq_4^4 x$ is no less coherent a proposition than $x=x$, and he is therefore left with no choice but to accept the consequences of this view. Fortunately, such revisions of deeply entrenched principles of logic---principles that, it may reasonably be conjectured, underlie all inquiry (including, ironically, the very inquiries that led Salmon to repudiate them)---prove to be quite unnecessary, given the hum-drum fact that, to know what is semantically encoded in the words they use, people must sometimes exploit information that is, in and of itself, semantic.

VIII. Modality

In these post-Kripke (1980) days, arguments like the following (in boldface) are widely seen as cogent:

(EP) Given only that the proposition *Hesperus is lovely* is true, it is nonetheless *epistemically* possible that the proposition *Phosphorous is not lovely* is true.

A closely related claim is that:

(NE) The proposition *Hesperus is lovely* doesn't entail the proposition *Phosphorous is lovely*.

At the same time, given that Hesperus *is* Phosphorous, it is impossible that one could be lovely if the other were not. So while it is *epistemically* possible that *Hesperus is lovely* should differ in truth-value from *Phosphorus is lovely*---in other words, while it is possible *given only what we know*---it is not *actually* or, as it is usually put, *metaphysically* possible. Thus:

(MI) given the truth of *Hesperus is lovely*, it isn't *metaphysically* possible that *Phosphorous is lovely* should be false (even though it is epistemically possible).

Thus, there is a distinction between epistemic and metaphysical possibility and, therefore, between epistemic and metaphysical necessity.

I wish to show that, in light of some remarks earlier made, we have reason to question the cogency of this argument and, more specifically, to question at least some (though not necessarily all) of the reasons adduced in favor of the view that metaphysical and epistemic necessity do not coincide. Consider the *sentences* (as opposed to the propositions):

(63) "Hesperus is lovely."
(64) "Phosphorous is lovely."

There is some object O (such that O is identical with the planet Venus)

such that “Hesperus” and “Phosphorous” both label O and such that those expression do nothing but label O. (So, at the level of semantics, those expressions do not bear Fregean senses and they aren’t Russellian quantifiers.) In terms of their semantics, (63) and (64) must therefore be synonymous, as each must have for its literal meaning the proposition:

(65) O is lovely.

Thus (EP) amounts to the following claim:

(EP*) given only that *O is lovely* is true, it is *epistemically* possible that *O is not lovely* is true.

But (EP*) is obviously false.

We’ve seen that, in virtue of having the form “…Hesperus…”, sentences typically *communicate* (as opposed to semantically encode) existence claims, the same being true of sentences having the form “…Phosphorous…” And we’ve also seen that what those existence-claims are is a function, not only of what these terms refer to, but also of what the circumstances are in which one is told what they refer to. It follows, as we have seen, that “Phosphorous is lovely” might communicate some proposition along the lines of:

(66) there is some object x such that x is uniquely a last celestial body to disappear from the morning sky; moreover, x is lovely,

and that “Hesperus is not lovely” might communicate some proposition along the lines of:

(67) there is some object y such that y is uniquely a first celestial body to appear in the evening sky; moreover, y is not lovely.

Obviously it is epistemically---and also, for that matter, metaphysically---possible that (66) and (67) should both be true. Neither proposition entails, or otherwise necessitates the truth of, the other.

Let us now conjoin (66) and (67):

(68) there is some object x such that x is uniquely a last celestial body to disappear from the morning sky; moreover, x is lovely; and there is some object y such that y is uniquely a first celestial body to appear in the evening sky; moreover, y is not lovely.

The following is obviously true:

(69) it is possible that there is some object x such that x is uniquely a last celestial body to disappear from the morning sky; moreover, x is lovely; and there is some object y such that y is uniquely a first celestial body to appear in the evening sky; moreover, y is not lovely.

(Of course, (68) is identical with the underlined part of (69).) Bearing all of this in mind, consider the following proposition:

(70) there is some object x such that x is uniquely a last celestial body to disappear from the morning sky; moreover, x is lovely; and there is some object y such that y is uniquely a first celestial body to appear in the evening sky; moreover, y is not lovely; moreover, it is impossible for the proposition *x is lovely* to differ in respect of its truth-value from the proposition *y is not*

lovely.

Before proceeding, let us clearly state some of our background assumptions. First, we're supposing that some *one* entity is uniquely a last celestial body to disappear from the morning sky and is also uniquely a first celestial body to appear in the evening sky. Second, we're supposing that nothing can simultaneously have and fail to have a given property. Third, we're setting aside subtleties relating to the fact that a given thing can be lovely at one time and non-lovely at some other time: we're assuming, for simplicity's sake, that a thing is always lovely if it's ever lovely. Given these assumptions, it follows that (70), taken as a whole, is false. (Why does it follow? Because, given those assumptions, a given thing would have to have and simultaneously fail to have a given property---that of being lovely---in order for (70) to be true.)

But what's important here is not that, as a whole, (70) is false. What's important here is that *the underlined part of (70) is true*. But---and this is equally important---notice that *there is no longer any proposition that is both necessary and epistemically contingent (or* a posteriori*)*. The underlined part of (69) is possible in both the epistemic and the metaphysical senses. (Obviously there is nothing metaphysically impossible about a world where something x is uniquely a celestial body to appear in the evening-sky, and where something y, distinct from x, is a last celestial body to disappear from the morning sky, and where x is lovely and y is not.) And the underlined part of (70), though necessary, *is also analytic*. There is some one object z (such that z is identical with Venus) such that the underlined part of (70) is true just in case the proposition *z is lovely* cannot possibly differ in truth-value from the proposition *z is lovely*. And, of course, it is analytic, for any object z, that *z is lovely* has the same truth-value as *z is lovely*.

What we are seeing is that, in at least some cases, there appears to be a distinction between epistemic and metaphysical necessity *only* if one fails to distinguish between literal meaning, on the one hand, and the information through which it is identified, on the other. For some O, what is literally meant by "Hesperus is lovely" coincides with what is literally meant by "Phosphorous is lovely." For some O, both sentences mean: *O is lovely.* But it is obviously *analytic* that, if *O is lovely* is true, so is *O is lovely.* So while it is indeed necessary that the proposition meant by "Hesperus is lovely" is true, given the truth of the proposition meant by "Phosphorous is lovely", that necessity is *analytic*: we aren't dealing with some kind of non-epistemic, *a posteriori* necessity.

At the same time, it is indeed true that a person could rationally assent to "Hesperus is lovely" while not assenting to "Phosphorous is lovely." But this has nothing to do with *a posteriori* necessary truth, and has to do instead with *a posteriori* non-necessary truth. (66) is the proposition *through* which one assigns meaning to "Hesperus is lovely", and (67) is the proposition *through* which one assigns meaning to "Phosphorous is lovely." The relationship between (66) and (67) is indeed *a posteriori*: given only that the one is true, it is an open empirical question whether the other is true. But not only is that relationship *a posteriori*: it is also *contingent*, as we discussed.

We therefore don't have anything that is *both* necessary *and* a posteriori. Given the truth of the proposition meant by "Hesperus is lovely", it is necessary that the proposition meant by "Phosphorous is lovely" is also true. *But that necessity is analytic*, the reason being that both sentences encode the same proposition. At the same time, given only an acceptance of "Hesperus is lovely", a person may need additional *empirical* information to determine whether "Phosphorous is lovely" is true. The reason is that, given that the proposition through which one identifies the literal meaning of the one is

true, it is not analytic (or otherwise epistemically necessary) that the proposition through which one identifies the literal meaning of the other is also true. But not only is not *analytic*: it is not necessary *at all*. Given the truth of the one proposition, it is *metaphysically* (and not just epistemically) possible that the other is false. The truth of (66) is metaphysically (as well as epistemically) compatible with the falsity of (67). Thus, we have analytic necessities and *a posteriori* contingencies---but no *a posteriori* necessities.

Another example may help clarify this argument. Consider the sentence:

(71) “Hesperus is identical with Phosphorous.”

For reasons that we’ve discussed, the existence-claim through which one accesses the proposition encoded in (71) is along the lines of:

(72) there is some object x such that x is uniquely a first celestial body to appear in the evening sky; and there is some object y such that y is uniquely a last celestial body to disappear from the morning sky; and, moreover, x=y.

Now consider the sentence:

(73) “it is necessarily the case that Hesperus is identical with Phosphorous.”

(Of course, (71) is identical with the underlined part of (73).) Given that (72) is the proposition that (71) conveys, it follows that (73) is ambiguous in respect of what *it* conveys, the reason being that the necessity-operator can be given different degrees of scope. The result of giving it narrow scope (or, more exactly, of giving it the narrowest possible scope) is:

(74) there is some object x such that x is uniquely a first celestial body to appear in the evening sky; and there is some object y such that y is uniquely a last celestial body to disappear from the morning sky; and, moreover, it is necessarily the case that x=y.

This is a true claim, and it captures the sense in which (71) is necessary while being non-trivial.

The result of giving the necessity-operator wide-scope (or, more exactly, of giving it the widest possible scope) is:

(75) It is necessarily the case that: there is some object x such that x is uniquely a first celestial body to appear in the evening sky; and there is some object y such that y is uniquely a last celestial body to disappear from the morning sky; and, moreover, x=y.

(75) is a false claim, since the underlined part is *not* necessary (either epistemically or metaphysically). What about the underlined part of (72)? There is no denying that it is necessary. After all, nothing could have not been itself. But is the underlined part of (72) epistemically contingent?

No. We must take care not to let facts about typography influence our views as to the constituencies of propositions. In particular, we mustn't be led astray by the fact that the expressions flanking the identity-sign in (72) are different. And we mustn't be led astray by the information to the left of the underlined part – we mustn't telescope that information into the proposition corresponding to the underlined part. Let P be the proposition corresponding to the underlined part. There is some one object z such that P is true just in case z=z and such that, to make a stronger claim, P is to the effect that z=z.

Of course, for any object z, the proposition *z=z* is metaphysically necessary (nothing could fail to be itself), but it is *also* analytic and therefore epistemically necessary.

There is some object x such that the literal meaning of (71) is the necessary – but also the analytic and, indeed, trivial – proposition *x=x.* When we say that (71) is *a posteriori* (i.e. epistemically contingent), we are talking, *not* about the proposition semantically encoded in that sentence, but *rather* about the proposition through which one ascertains the proposition semantically encoded in it. So the proposition that is necessary is quite distinct from the proposition that is epistemically contingent: we thus haven't found anything that is both necessary *and* a posteriori. And in saying that (71) is both necessary and *a posteriori*, one is guilty of failing to distinguish literal meaning from the information through which it is identified.

IX. Saul on substitution-failures in extensional contexts Jennifer Saul (1997) identified a problem on which the analysis defended in this paper may be capable of shedding some light. Like most semanticists, Saul assumes that intersubstituting co-referring terms can lead from true sentences to false ones and *vice versa.* (Even though I have argued against that assumption, I will assume its truth when summarizing Saul's article.) But, unlike most semanticists, Saul *denies* that substitution-failures occur only when an expression occurring in a non-extensional context is replaced with a co-referring expressions. Saul provides excellent *prima facie reasons* for holding this.

Here is Saul's argument. Consider the following sentences:

(76) "Clark Kent went into a phone booth and Superman came out."
(77) "Clark Kent went into a phone booth and Clark Kent came out."

(76) and (77) are---or, in any case, very much seem to be---capable of differing in truth-value, even though neither “Clark Kent” nor “Superman” occurs non-extensionally in either of them. This suggests that extensional contexts are sometimes substitution-intolerant.

This is a serious problem. Saul considers four attempts to solve this problem. She rejects each of them, and her reasons for doing so are cogent. (In Section X, we’ll consider the solutions that Saul rejects along with her reasons for rejecting them.) But Saul doesn't put forth a solution of her own. Fortunately, since the analysis defended in this paper practically *is* just a solution, we are in a position to step into this breach.

A story will help make this clear. While sitting on a park bench, you and I see a person in the distance. We don't know who that person is. And although we obviously have *some* idea of what this person looks like, our knowledge of his or her appearance is extremely vague. Being so far away, he or she is little more than a vague blur. There is only one non-generic fact about this person of which our blurry perceptions *do* apprise us: he or she is wearing brightly colored, loose fitting clothes.

For whatever reason, we wish to talk about this person. To make it easier to talk about him or her, we decide to give him or her a name. We decide to refer to this person as "Argo."

Because we're both semanticists, we don't want to put our neologism to use until we've precisely identified the semantic rule that is to govern its use (i.e. that states how, in virtue of that expression’s occurring in a given sentence, that sentence’s meaning and/or truth-conditions are thereby affected). After a few minutes of quibbling, we write down the following (the material in boldface), and agree to abide by it:

(AR) There is a unique thing O such that O is a contextually salient person in the distance such that (i) O is wearing brightly colored, loose-fitting clothes and such that (ii) we are hereby agreeing that (at least as far as conversations between the two of us are concerned) "Argo" is to refer directly to O.

In agreeing that "Argo" is to refer *directly* to the aforementioned person, we are agreeing that, in virtue of having the form

"...Argo...",

a statement S encodes a proposition P such that P is true exactly if

....O....

Thus, the remaining semantic rules of English being what they are,

"Argo is an economist"

encodes a proposition P that is true exactly if:

O is an economist. Thus, for "Argo is an economist" to be true, it is necessary and sufficient that O be an economist, it being irrelevant what other properties O has. Given any statement S^ of the form

"....Argo...",

the same thing *mutatis mutandis* holds of S^; in other words, S^ encodes a proposition P that is true exactly if

O is a thing x such that...x...

Having agreed that AR states the semantic rule for "Argo", you and I then continue to watch Argo. Argo starts doing extraordinarily difficult acrobatics. He continues to do them for hours; he's clearly a master-gymnast.

We spend several hours watching him perform these stunning feats. During that time, Argo is, or at least seems to be, unaware of our presence. But eventually Argo clearly notices us. Argo doesn't approach us; nor we do approach him. So our knowledge of what he looks like doesn't improve, and neither does our knowledge of who he is. After locking eyes with us, Argo resumes his gymnastics.

Twenty minutes pass. Argo suddenly stops his gymnastics. Looking right at us, Argo shouts: "Hi! Nice to meet you. I am an investment banker, but I like to do acrobatics in my spare time."

Argo's voice is clearly that of a man. And it's distinctive in a number of other respects. It's the voice of somebody who is healthy---of somebody who doesn't have emphysema or any other ailments that might affect the timbre or resonance of a person’s voice. Also, Argo's voice seems to be that of somebody who is trained at public speaking; and, partly for that reason, though (it seems) not entirely, it exudes both confidence and good will. Having said these few words, Argo suddenly leaves.

The next day, you and I are sitting on the same park bench. Once again, we see a person in the distance. We know nothing about this person. And, being as far from us as Argo was, our knowledge of his or her appearance is as vague as our knowledge of Argo's appearance. There is only one non-generic fact about this person of which our blurry perceptions *do* apprise us: he or she is wearing drab, tight-fitting clothes.

Wishing to speak about this person, but not knowing his or her name, we decide to refer to him or her as "Ogra." Being semanticists, we decide, before using our neologism, to identify the semantic rule that is to govern its usage. We agree that the semantic rule for "Ogra" is given by a statement, which we'll call "OG", that is exactly like AR *except* that (a) OG contains an occurrence of "Ogra" in the place where AR contains an occurrence of "Argo" and (b) OG contains an occurrence of "O is wearing drab, tight-fitting clothes" in the place where AR contains an occurrence of "O is wearing brightly colored, loose-fitting clothes."

Having agreed that OG states the semantic rule for "Ogra", we continue to watch Ogra. Ogra is very clumsy, or at least seems to be. Ogra tries to do a summersault but fails. He or she then tries to do a cart-wheel, but once again fails. Moreover, Ogra's gait, when not doing gymnastics, is clumsy. And Ogra's posture is poor; he or she slouches. If Ogra doesn't yet suffer from scoliosis, he or she soon will, in all likelihood. Argo's posture, by contrast, was exemplary.

Ogra continues his or her oafish attempts at gymnastics for several hours. During this time, he or she doesn't notice us. But then she stares right at us, clearly aware that we've been watching himher. Ogra shouts out: "I wish you'd let me known that you were here. I feel terribly embarrassed. By the way, I am investment banker." Though clearly that of a man, Ogra's voice is otherwise very different from Argo's. Ogra's voice is weak and raspy. He wheezes heavily, and his statement, though short, is interrupted by hacking coughs. Moreover, the tone of Ogra's voice was that of a resentful, weak, sour person. After his plaintiff valediction to us, Ogra leaves.

Day after day, we return to the same park bench. On each day, either Ogra or Argo appears. On some days they both appear; and sometimes each appears on several occasions. But they never appear at the same time.

Every time either appears, each behaves just as he did the first time we saw him. (For example, whereas Agro always performs brilliant acrobatic feats, Ogra always tries, without success, to perform rudimentary acrobatic feats. And, whereas Argo's valedictions exude vigor and self-confidence, Ogra's exude weakness and insecurity. And so on.) Moreover, nothing happens, and nobody tries to make anything happen, that would enable either of us to learn anything new about either of them. Neither of them approaches either of us or otherwise does anything that would tip of us as to his identity, and neither of us approaches either of them or otherwise does anything to learn more about his identity. Thus, neither of us ever learns anything new about either of them.

Our story about Argo and Ogra isn't over yet; but, before we can complete it, we must a few points clear. There is some object O such that the proposition literally meant by the statement:

(78) "Argo is an investment banker"

is true exactly if:

(79) O is an investment banker,

it being entirely irrelevant, as far as semantics (literal meaning) is concerned, what other properties O might have.

But, the circumstances being what they are, if I utter (78) to you, the message (or, in any case, part of the message) that will thereby be conveyed, and that (so we may plausibly suppose) it is my intention to convey, is along the lines of:

(80) There is somebody x such that we've been in the habit of watching x skillfully perform difficult acrobatics and such that x's voice has a healthy resonance; moreover, x is an investment banker.

And, when speaking to you, were I to produce an utterance of:

(81) "Ogra is an investment banker",

the same points *mutatis mutandis* would hold of that utterance, so that, for some object O*, the proposition literally meant by that utterance would be:

(82) O* is an investment banker,

even though the proposition *conveyed* by such an utterance would coincide *mutatis mutandis* with (80).

Let us continue our story. For some reason, there is a phone booth in the park in question. During each of our now habitual Argo-watches, we notice that Argo goes into the phone booth and spends around five minutes in there. (Presumably he's talking on the phone. But it doesn't matter whether that presumption is correct.) Given AR (the semantic rule governing "Argo"), it follows that, there is some object O (namely, Argo) such that the (true) proposition P literally meant by an utterance of:

(83) "Argo went into that phone booth around five minutes ago, and Argo exited it around ten seconds ago"

is equivalent (if not identical) with:

(84) O went into that phone booth around five minutes ago, and O exited it around ten seconds ago.

But obvious applications of now familiar reasoning show that, the circumstances being what they are, were I, during one of our Argo-afternoons, to utter (83) to you, the message (or, in any case, part of the message) that would thereby be conveyed, and that (so we may plausibly suppose) it would my intention to convey, would be along the lines of:

(85) There is somebody x such that we've been in the habit of watching x skillfully perform difficult acrobatics and such that x's voice has a healthy resonance; moreover, x went into that phone booth around five minutes ago, and he exited it around ten seconds ago.

Exactly similar points hold of utterances of:

(86) "Ogra went into that phone booth around five minutes ago, and Ogra exited it around ten seconds ago."

Given this, suppose that, while addressing you under the circumstances in question, I were to utter (86). It follows from what we've said that (a part of) what was thereby conveyed to you would be along the lines of:

(87) There is somebody y such that we've been in the habit of watching y oafishly try to perform rudimentary acrobatics and such that y's voice has a unhealthy, plaintiff resonance; moreover, y went into that phone booth around five minutes ago, and he exited it around ten seconds ago.

(85) and (87) are distinct propositions, and neither entails the other. In fact, given the circumstances, it would be stretch to say that either so much as confirms the other.

Bearing all of this in mind, consider the following sentence:

(88) “Argo went into that phone booth around five minutes ago, and Ogra exited it around ten seconds ago.”

There is some object O (namely, Argo), and some object O* (namely, Ogra), such that the proposition literally meant by an utterance of (88) is:

(89) O went into that phone booth around five minutes ago, and O* exited it around ten seconds ago.

But, the circumstances being what they are, were I to utter (88) to you, the proposition thereby communicated would be along the lines of:

(90) There is somebody x such that we’ve been in the habit of watching x skillfully perform difficult acrobatics and such that x’s voice has a healthy resonance; moreover, x went into that phone booth around five minutes ago; and there is somebody y such that we’ve been in the habit of watching y oafishly try to perform rudimentary acrobatics and such that y’s voice has a unhealthy, plaintiff resonance; moreover, y exited that phone booth around ten seconds ago.

Note that each of (85), (87), and (89), fails to entail, or to be entailed by, either of the other two.

Let us continue our soon to be completed story. Each of Argo and Ogra has

claimed to be an investment banker. But that, in our opinion, is where the similarities between them end. True—they've never both appeared at the same time. But, given the remaining data, that fact provides little or no weight for the hypothesis that Argo is Ogra.

Now for the conclusion of our story. One day Argo walks up to us and says to us "it's good to meet you, at long last." His voice is exactly what we'd expect it to be. Argo then says "I'm sure you have many questions about me." But, in uttering *this* sentence, his voice is clearly Ogra's. What we are hearing isn't merely a good impersonation of Ogra's voice: it *is* Ogra's voice. (Let's suppose that, for some reason, you and I are voice-experts who can distinguish excellent impersonations of any given voice from that voice itself.) Argo then strips off his loose fitting, brightly colored vestments. Underneath them are the drab, tight-fitting clothes that Ogra wears. Argo then does various summersaults and cartwheels and he does them in *exactly* the way in which Ogra does them. It's clear that Argo isn't *impersonating* Ogra. Argo *is* Ogra.

Given the semantic rules governing "Argo" and "Ogra"---AR and OG, in other words---it follows that there is some one object O such that an utterance of:

(91) "Argo is Ogra"

has for its literal meaning a proposition:

(92) O is O.

But it's also clear that, given the circumstances, an utterance of (91) would convey (inter alia) a proposition that, unlike (92), is non-trivial and, indeed, is

extremely rich in empirical information, that proposition being along the lines of:

(93) There is somebody x such that we've been in the habit of watching x skillfully perform difficult acrobatics and such that x's voice has a healthy resonance; and there is somebody y such that we've been in the habit of watching y oafishly try to perform rudimentary acrobatics and such that y's voice has a unhealthy, plaintiff resonance; moreover, x is identical with y.

It also follows that (83), (86), and (88) *all have the same literal meaning.* We've already made it clear explained how it is that, despite that fact, they convey such different propositions.

Everything we said about the sentence-pair consisting of (83) and (86), and about the sentence-pair consisting of (86) and (88), holds *mutatis mutandis* of the sentence-pair consisting of (76) and (77). Also, (79) and (82) are, it turns out, one and the same proposition, notwithstanding that, for now familiar reasons, the proposition conveyed by an utterance of (78) could be very different *ceteris paribus* from the proposition conveyed by an utterance of (81).

X. Saul on substitution-failures in extensional contexts (continued) Even though Saul doesn't put forth a positive analysis of such sentence-pairs, what she does say is consistent with, and indeed favorable to, the analysis we have defended in this paper. Saul rejects analyses that are incompatible with ours; and, without going so far as to endorse it, she says of an analysis that is consistent with ours, and embodies the conceit lying at the center of it, that it isn't as unworthy of consideration as semanticists are wont to believe. Let's now consider what Saul says in some detail.

According to Frege,

(94) “Smith thinks that Clark Kent went into a phone booth and that Clark Kent came out”

may differ in truth-value from

(95) “Smith thinks that Clark Kent went into a phone booth and that Superman came out.”

For argument’s sake, let’s suppose that Frege is right about this. Frege’s way of dealing with this substitution-failure is to say that, in (94), the occurrences of “Clark Kent” refer to the *sense* that “Clark Kent” ordinarily (i.e. in substitution-tolerant contexts) bears, the same thing *mutatis mutandis* being true of the occurrence in (95) of “Superman.” Frege held that all substitution-failures are to be explained along similar lines.

But, Saul rightly says, the occurrences in (76) and (77) of “Clark Kent” and “Superman” don’t refer to senses. They refer to people (or, rather, to *a* person). So Frege’s solution is a non-starter, at least in this context.

Like Frege, Ted Cider (1996) takes it for granted that, at the level of literal meaning, (76) and (77) differ in truth-value. But Cider’s way of dealing with this is non-Fregean. According to Cider, the reason that (76) and (77) differ in truth-value is that “Superman” and “Clark Kent” refer to different *time-slices* of some one person. So the occurrences in (77) of “Clark Kent” and “Superman” do not co-refer, according to Ted Cider; they refer to distinct phases of some one entity.

Saul makes it clear that Cider’s analysis cannot be accepted. A consequence of Cider’s view that “Clark Kent” and “Superman” don’t co-

refer is that “Superman is identical with Clark Kent” is false. But since that sentence is true, Cider’s analysis is false.

Cider, as Saul points out, might respond by saying either that “is” is ambiguous---that it sometimes denotes the relation of identity and sometimes denotes some other relation---*or*, as Saul also points out, Cider might respond by saying that “Clark Kent” and “Superman” don’t have fixed references and that, for that reason, they co-refer in some contexts (e.g. “Superman and Clark Kent are one and the same”) but don’t co-refer in others (e.g. (77)). But, as Saul says, such positions, in addition to being *ad hoc*, are demonstrably unacceptable, the reason being that they fail to deal with sentences such as (to use a variant of Saul’s own example) the following:

(96) “Clark Kent often flies to work, though he conceals this fact.”

If Cider is right, the thing denoted by “Clark Kent” is a time-slice that *doesn’t* fly. Moreover, as Saul says, since neither the word “is”, nor any variant thereof (e.g. “is identical with”), appears in (96), no amount of tinkering with the semantics of such expressions is of any use.

Saul doesn’t consider what Russell would say about the (apparent) fact that (76) differs in truth-value from (77). But this doesn’t much matter, since Russell’s view is false, as Kripke (1977, 1980) made clear.

We’ve consider all of the analyses that Saul rejects. There is one analysis that she considers that she doesn’t reject, even though, as previously stated, she doesn’t go so far as to accept it. According to Salmon (1986) and Soames (2002), (76) and (77) *do* have the same truth-value and, indeed, the same literal meaning. They grant that (76) and (77) *seem* to differ in truth-value. But this, they say, is because pragmatics is, in some way or other, obscuring semantics.

Why doesn't Saul embrace this solution? Because in the form in which Salmon and Soames present it, Saul rightly says, it leaves too many crucial questions unanswered. Salmon and Soames say that *in some way or other* pragmatics is warping our views concerning literal meaning. But they don't say *how* it is doing so, even though, given their position, it is incumbent on them to provide this information. And—to develop this point---Salmon and Soames don't answer either of the following questions, even though it is incumbent on them to answer both:

(i) What are the messages that are (pragmatically, non-semantically) *communicated* by (tokens of) (76) and (77)? (To generalize this question: given a sentence having either the form "…Clark Kent…" or the form "…Superman…", what is the proposition that, in virtue of having that form, utterances of that sentence convey? To generalize this question even further: given two distinct, co-referring proper N and N*, and given a sentence having either the form "…T…" or the form "…T*…", what is the proposition that, in virtue of having that form, utterances of that sentence convey?)

(ii) Supposing for argument's sake that t and t* are tokens of (76) and (77), respectively, and that T and T* (non-semantically) impart propositions P and P*, *why* do those tokens convey those propositions? What is the principle at work? (To generalize this question: given a sentence-token having either the form "…Clark Kent…" or the form "…Superman…", what is the proposition that, in virtue of having that form, that token conveys? To generalize this question even further: given two distinct, co-referring proper N and N*, and given sentence-tokens T and T* that, in virtue of having the forms "…N…" "…N*…", respectively, convey propositions P and P*, *why* is it that, in

virtue of having those forms, T and T* communicate P and P* specifically? Why those propositions, as opposed to others?)

As Saul points out, many recoil from the position advocated by Salmon and Soames. Suppose that Clark Kent (i.e. somebody x such that x is Clark Kent/Superman and such that x looks and acts like a bungling, bespectacled reporter, etc.) goes into a phone booth and that Superman (i.e. somebody y such that y is identical with Superman/Clark Kent and such that y looks and acts like a caped crime-fighter who has superhuman abilities, etc.). Under these circumstances, (76) would, whereas (77) would not, be seen as a correct (or, in any case, non-misleading) statement. But given that

(97) "Clark Kent went into a phone booth and Clark Kent came out, but nobody recognized him"

is *not* a misleading characterization of what happened, it follows that the message borne by (76), so far as it's a misleading one, can be cancelled. Since it can be cancelled, it is no part of what (77) literally means.

This insight of Saul's provides additional corroboration for our contention that, once the phenomenon of presemantic implicature is taken into account, the relevant linguistic intuitions *are* consistent with the view that (76)'s literal meaning coincides with (77)'s literal meaning and, more generally, with the contention that proper names are directly referential.

XI. Boguslawski's analysis

Like us, Andrzej Boguslawski (1994) holds that intersubstituting co-referring expressions preserves truth-value. Moreover, Boguslawski's

argument for this contention is similar to our argument for it. Boguslawski's argument can be stated in terms of the following piece of fiction.

Jones lives next door to Smith. Smith knows this, and Smith also knows that Jones is a talented pianist. But unbeknownst to Smith, Jones is also the tallest American spy. In light of this, consider the following three sentences:

(98) Smith knows *of* the person who is in fact the tallest American spy that he is a talented pianist.
(99) Smith knows that his neighbor is a talented pianist.
(100) Smith knows that the tallest American spy is a talented pianist.

Of course, supposing that definite descriptions are referring terms, the underlined expressions all co-refer. (In this section, it will be assumed, for argument's sake, that definite descriptions are indeed referring terms.)

It is uncontroversial that (98) and (99) are correct. But it appears, at least at first, that (100) is false. If (100) is false, then it obviously follows that not all contexts are extensional. But the following extension of our story suggests that, despite first appearances, (100) is in fact correct.

In Smith's absence and without his knowledge, I am now talking to Jones. I say to Jones:

(101) Smith knows that you are a talented pianist.

Under the circumstances, (101) strongly appears to be correct. At the same time, Smith has no idea that I am now talking to Jones, and he therefore has no idea that I am referring to Jones with the word "you." For somebody to be correctly referred to with the term "you" in context C, that person must

satisfy some description along the lines of: *person x such that x is being addressed in C*. So even though (101) is plainly correct, Smith nonetheless has no idea that Jones satisfies the descriptive information associated with the expression that, in (101), is used to pick Jones out. Thus, there is some person x (namely, Jones) such that the following three conditions are met:

(a) the occurrence of "you" in (101) refers to x;
(b) in the situation in question, x uniquely satisfies the description *person to whom I am speaking*; and
(c) Smith does not *know* that x satisfies that description.

Thus, the truth of (101) seems to depend only on whether, for the relevant value of x, Jones knows that *x* is a talented pianist, it being irrelevant what kind of descriptive information is used to refer to x. This line of thought is consistent with (even though, on its own, it admittedly falls far short of establishing) extensionality.

A continuation of our story provides additional corroboration for this principle. I am now talking with Green. Jones is standing near us, and is wearing a blue shirt. Nobody besides Jones is wearing a blue shirt. I point to Jones and say:

(102) Smith knows that <u>the guy wearing a blue shirt</u> is a talented pianist.

Under the circumstances, (102) seems to be correct. At the same time, Smith may have absolutely no idea that Jones is now wearing a blue shirt. Thus,

there is some person x (namely Jones) such that the following three conditions are met:

(a*) the occurrence of “the guy wearing a blue shirt” in (102) refers to x;
(b*), in the situation in question, x uniquely satisfies the description *person wearing a blue shirt* and
(c*) Smith does not *know* that x satisfies that description.

Thus, the truth of (102) seems to depend only on whether, for the relevant value of x, Jones knows that *x* is a talented pianist, it being irrelevant what kind of descriptive information is used to refer to x.

The following variation of the story just told provides yet further support for extensionality. I am talking with Green. Jones is visible to both of us, and he has just kicked a soccer ball. Green doesn’t know Jones’ name, but Green does know Smith’s name. I say to Green:

(103) Smith knows that <u>the guy who just kicked a soccer ball</u> is a talented pianist.

Under the circumstances, (103) seems to be correct. It seems irrelevant whether Smith knows that Jones has just kicked a soccer ball. All that matters is that, for the relevant value of x, Smith knows that x is a talented pianist.

The story just told can easily be developed so as to show that, given the truth of (98) and (99), the truth of (100) follows. Green and I are both spies. We are both acquainted with Jones, and we both know that he is the tallest American spy. But, for some reason, we both keep on forgetting his name. Under these circumstances, if, addressing Green, I utter (100), the resulting

token of (100) *does* very much seem to have the same truth-value as (99). A closely related point is that, under these circumstances, there is some object O such that O is identical with Jones, and such that each of (99) and (100) affirms the singular proposition:

(104) Smith knows that O is a talented pianist.

There is no denying that, at least some occasions, (99) and (100) *communicate* propositions that differ in truth-value. But this is easily explained without supposing that what is literally meant by a token of the one differs from what is literally meant by a token of the other. For reasons that we've already discussed at length, a consequence of the direct-referentialist's position is that:

(DDI) In virtue of having the form, "…the phi…", a sentence S will implicate the proposition: *exactly one thing x has phi and…x…*

Given (DDI), there are two possibilities as to what (99) will implicate. It could implicate either:

(99_{NS}) Smith knows that exactly one thing x is his neighbor and, moreover, that x is a talented pianist,

or

(99_{WS}) exactly one thing x is Smith's neighbor; moreover, Smith knows that x is a talented pianist.

The implicature can be given either wide scope or narrow scope; and what degree of scope it is given is to be understood in terms of standard principles of pragmatics, as a bit of fiction shows. Both Brown and Jones know that, for some x, x is Smith's neighbor and is also a talented pianist, and each knows the other to know this. Further, each of Brown and Jones knows the other to be a highly rational person. Brown says to Jones:

(105) "Smith knows that his neighbor is a talented pianist; and Smith knows that the tallest American spy is a talented pianist."

Notice that (105) contains an occurrence of (99) and also of (100). If the corresponding implicatures are given wide-scope, then Brown is inexplicably repeating himself; for, thus interpreted, there is some x such that Brown is saying: *Smith knows that x is a talented pianist and Smith knows that x is a talented pianist.* But Brown's utterance becomes perfectly intelligible if it is supposed that at least one of those implicatures is to be given narrow-scope; for, thus interpreted, Brown's utterance does not consist of a sterile repetition of some one assertion.

To sum up: given (DDI) along with some plausible views concerning pragmatics, it needn't be supposed that truth-value can be changed by replacing a definite description with a co-referring description.

But supposing, if only for the sake of argument, that such intersubstitutions *can* change truth-value, extensionality is still not threatened in any way. Given that Smith's one neighbor is the tallest American spy, (99) and (100) can diverge in truth-value only if they are equivalent with (99_{NS}) and (99_{WS}), respectively. But in that case, for the reasons discussed earlier, "his neighbor" and "tallest American spy" are quantifiers, not referring terms and therefore not *co*-referring terms, a corollary being that any difference in

truth-value between (99) and (100) has nothing to do with any substitution-failure.

Conclusion

In their efforts to explain relatively innocuous facts relating to language (e.g. the fact, discovered by Kripke (1980), that proper names are not synonymous with definite descriptions), philosophers have advocated wholesale revisions of logic and logic (e.g. the contention, urged by Nathan Salmon (2007), that for some values of x, the information at a person's disposal may confirm the proposition that $x\overset{4}{4}x$ as much as it does the proposition that x=x). Fortunately, given the pedestrian fact that people must sometimes exploit non-linguistic information to make linguistic judgments, these revisions prove to be as unnecessary as they are grotesque.

References

Anderson, C. Anthony. 1986. Some Difficulties Concerning Russellian Intensional Logic. Noûs 20 (1), 35-43.

Austin, J.L. 1953. How to do things with words. Clarendon Press, Oxford.

Bach, Kent. 1987. Thought and Reference. Clarendon Press, Oxford.

Barwise, Jon and Perry, John. 1983. Semantic innocence and uncompromising situations. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 392-404.

Bezuidenhout, A., et al. (Eds.), Descriptions and Beyond. Oxford University Press, New York.

Boguslawski, Andrzej. 1994. Sentential Complementation and Truth. In: Čmejrková, S., et al. (Eds.), The Syntax of Sentence and Text. John Benjamins, Philadelphia, PA, pp. 317-332.

Cappelen, Herman and Lepore, Ernie. 2005. Insensitive Semantics: A defense of speech act pluralism and semantic minimalism. Blackwell, Oxford.

Chomsky, Noam. 1980. Rules and Representations. Columbia University Press, New York.

Chomsky, Noam. 1988. Language and the problems of knowledge. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 509-527.

Cider, Ted. All the world's a stage. Australasian Journal of Philosophy 74, 433-457.

Davidson, Donald. 1967. Truth and Meaning. Synthese 17, 304-323.

Davidson, Donald. 1968. On saying that. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 352-361.

Davidson, Donald. 1980. Inquiries into Truth and Interpretation. Oxford University Press, New York.

Donnellan, Keith. 1966. Reference and definite descriptions. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 235-247.

Dummett, Michael. 1973. Frege: Philosophy of Language. Harvard University Press, Cambridge, MA.

Evans, Gareth. 1982. The Varieties of Reference. Clarendon Press, Oxford.

Frege, Gottlob. 1892. On sense and nominatum. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 190-202.

Jackendoff, Ray. 1989. What is a concept, that people may grasp it? In: Laurence, E., et al. (Eds.), Concepts: Core Readings. The M.I.T. Press, Cambridge, MA, pp. 305-334.

Kaplan, David. 1968. Quantifying In. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 362-391.

Kaplan, David , 1975. Dthat. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 316-329.
Kaplan, David. 1975b. How to Russell a Frege-Church. Journal of Philosophy 64, 636-649.
Kaplan, David. 1979. On the Logic of Demonstratives. Journal of Philosophical Logic 8, 71-98.

Kaplan, David. 1986. Opacity. In: Hahn, L ., et al. (Eds.), The Philosophy of W.V.O. Quine. Open Court Press, LaSalle, IL, pp. 229-289.

Kaplan, David. 1989. Demonstratives. In: Almog, J., et al.(Eds.), Themes from Kaplan. Oxford University Press, New York, pp. 471-555.

Kaplan, David. 1989b. Afterthoughts. In: Almog, J., et al. (Eds.), Themes from Kaplan. Oxford University Press, New York, pp. 557-534.

Kripke, Saul. 1977. A Puzzle about Belief. In: Margalit, A. (Ed.), Meaning and Use. D. Reidel, Dordrecht.

Kripke, Saul. 1977b. Speaker's Reference and Semantic Reference. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 248-258.

Kripke, Saul. 1980. Naming and Necessity. Harvard University Press, Cambridge, MA.

Kuczynski, John-Michael. 2004. A non-Russellian treatment of the referential-attributive distinction. Pragmatics and Cognition 12 (2), 253-294.

Kuczynski, John-Michael. 2006. On Marga Reimer's Descriptions and Beyond. Pragmatics and Cognition, 14 (1), 196-204.

Kuczynski, John-Michael. 2007. Conceptual Atomism and the Computational Theory of Mind: a Defense of Content-internalism and Semantic externalism. John Benjamins, Amsterdam.

Kuczynski, John-Michael. 2009. Analytic Philosophy. Kendall Hunt, Dubuque, IN.

McCawley, James D. 1971. Everything that Linguists Have Always Wanted to Know about Logic. University of Chicago Press, Chicago, IL.

Moore, Robert C. 1995. Logic and Representation. CSLI Publications, Palo Alto, CA.

Neale, Stephen. 1990. Descriptions. The M.I.T. Press, Cambridge, MA.

Parsons, Terence. 2000. Indeterminate Identity. Oxford, New York.

Peacocke, Christopher. 1992. A Study of Concepts. The M.I.T. Press, Cambridge, MA.

Perry, John. 2000. The Problem of the Essential Indexical and Other Essays. CSLI Publications, Palo Alto, CA.

Pinker, Steven. 2007. The Stuff of Thought: Language as a Window Into Human Nature. Viking Books, New York.

Priest, Graham. 2006. Doubt Truth to be a Liar. Oxford, New York.

Quine, Willard van Orman. 1953. Reference and Modality. In: Quine, W.V.O. (Ed.), From a Logical Point of View. Harvard University Press, Cambridge, MA, pp. 139-151.

Quine, Willard van Orman. 1960. Word and Object. The M.I.T. Press, Cambridge, MA.

Recanati, Francois. 1989. The pragmatics of what is said. Mind and Language 4, 295-329.

Recanati, Francois. 1993. Direct Reference: From Language to Thought. Blackwell, Oxford.

Recanati, Francois. 2004. Literal Meaning. Cambridge University Press, Cambridge.

Russell, Bertrand. 1905. On Denoting. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 203-211.

Russell, Bertrand. 1918. The philosophy of logical atomism. In: Marsh, R.C. (Ed.), Logic and Knowledge. Routledge, London, pp. 167-272.

Salmon, Nathan. 1986. Frege's Puzzle. The M.I.T. Press, Cambridge, MA.

Salmon, Nathan. 2005. Metaphysics, Mathematics, and Meaning: Philosophical Papers I. Oxford University Press, New York.

Salmon, Nathan. 2007. Content, Cognition, and Communication: Philosophical Papers II. Oxford University Press, New York.

Saul, Jennifer. 1997. Substitution and simple sentences. Analysis 57/2, 102-

108.

Searle, John. 1970. Expression and Meaning. Cambridge University Press, Cambridge.

Soames, Scott. 2002. Beyond Rigidity: The Unfinished Semantic Agenda of *Naming and Necessity*. Oxford University Press, New York.

Smullyan, Raymond. 1963. Modality and description. In: Linsky, L. (Ed.), Reference and Modality. Oxford University Press, New York, pp. 35-43.

Strawson, Peter. 1950. On Referring. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 219-234.

Tarski, Alfred. 1973. The concept of truth in formalized languages. In: Corcoran, J. (Ed.), Logic, Semantics, and Metamathematics. Hackett, Indianapolis, IN, pp. 152-169.

Travis, Charles. 1975. On what is strictly speaking true. Canadian Journal of Philosophy 15 (2), 187-229.

Tughendat, Ernst. 1970. The meaning of “Bedeutung” in Frege. Analysis 30, 168-189.

Wittgenstein, Ludwig. 1958. Philosophical Investigations. Blackwell, Oxford.

Wreen, Michael. 1974. Belief, modality, opacity, and the

referential/attributive distinction. Linguistics 22, 313-340.

Analyticity

Here is a view that was once widely held.

Triangles have three sides is a priori, and so is *all cases of knowledge are cases of belief.* The truth of these propositions is to be known through examination of concepts (in the platonic, non-psychological sense), not through examination of the spatio-temporal world.

The corresponding sentences – 'triangles have three sides' and 'all cases of knowledge are cases of belief' – are characterized as *analytic.*

If a sentence bears an a priori proposition, then it is analytic.[360] The sentence 'Socrates was bald' is not true *wholly* in virtue of the meaning that it bears; given *only* that 'Socrates is bald' means *Socrates is bald*, it by no means follows that 'Socrates is bald' is true.

But the story is different with 'triangles have three sides.' Given only that 'triangles have three sides' means *triangles have three sides*, it follows that 'triangles have three sides' is true. That sentence is true wholly in virtue of what it means. Some sentences are true wholly in virtue of meaning.

Until recently, virtually every philosopher, apart from a few empiricist hardliners, accepted that there were a priori truths and thus a priori knowledge. Thus, it was accepted that there were analytic sentences –

sentences that hold entirely in virtue of their meanings. So radical empiricism – the thesis that *all* knowledge is a posteriori -- was generally rejected.

Of course, if a proposition is *a priori* true, then it cannot rationally be given up in light of experience. So if there are a priori propositions, then some propositions must have a fixed place in our belief-system. The same is true of the corresponding sentences (i.e. those sentences cannot be given up, at least not without assigning them different meanings).

The views just described are now widely considered false. This is due mainly to Quine's 'Two Dogmas of Empiricism'. Now many accept that there are no a priori truths and no analytic sentences. *Squares have four sides* could, in principle, be disconfirmed by experience. The same is true, therefore, of the corresponding sentence.

In fact, thanks to Quine, the non-existence of a priori truth is practically seen as a datum by many, perhaps most, contemporary analytic philosophers. Paul Boghossian sums it up:

This is what many philosophers believe today about the analytic/synthetic distinction: In his classic early writings on analyticity…Quine showed that there can be no sentences that are true purely by virtue of their meaning, and those that aren't. In so doing, Quine devastated philosophical programs that depend upon a notion of analyticity – specifically the linguistic theory of necessary truth, and the analytic theory of a priori knowledge…I don't know precisely how many philosophers believe all of the above, but I think it would be fair to say that it was the prevailing view. [361]

I will argue two points. First, Quine's argument establishes nothing. Second, there are a priori truths, and thus sentences that express such truths (analyticities); there are non-denumerably many of them.

Why criticize Quine's argument? That argument embodies fundamental misconceptions about several important issues, including the nature of logical truth, the relation of semantics to syntax, the nature of meaning, the possibility of 'naturalizing' meaning, the concept of synonymy, and the nature of rationality. Studying Quine's argument is a good way to identify, and thus to begin correcting, many deeply entrenched, but very wrong, ideas about these matters.

I will present several arguments. Some of these were given, or at least anticipated, by Bonjour, Dummett, and Wright. But many of them are new to this paper. More accurately, although those points *per se* are not all of them new to this paper, their bearing on Quine's argument hasn't yet been recognized. So my anti-Quinean *use* of these points is new to this paper.

I must qualify a point made a moment ago. Quine doesn't say that there are *no* analyticities. But he holds that the only analyticities that exist are trivial. He concedes that 'all green squares are green' is analytic, that 'all unmarried males are males' is analytic. (How could he possibly *deny* an analytic status to such sentences?) If a sentence is *formally* true – is such that facts about its *syntactic structure* are enough to establish its truth – then Quine concedes that it is analytic.

But, for Quine, those are the *only* analytic sentences. So 'all bachelors are unmarried' is not analytic, and neither is 'all green squares are four-sided figures.' For such sentences are not *formally* true; they have the same syntax as 'all bachelors are grumpy' and 'all green squares are made of copper', which are not analytic.

We will see later how, in conceding an analytic status to 'all unmarried males are males', he makes it impossible to avoid conceding an analytic status to 'all bachelors are unmarried' and even "the interior angles of a triangles (in a Euclidean space) add up to 180°."

I don't give Quine's argument *itself* until section 3. Up till then, I criticize some of the foundational assumptions that underlie that argument. But readers eager to hear a discussion of Quine's argument *per se* may skip to that section.

2. First of all, Quine's ostensible target is analyticity, not a priority. Quine discusses whether there are *sentences* that cannot be given up in light of experience, not whether there are *propositions* that have that property.

But, as Dummett[362] points out, what Quine says about analyticity maps cleanly onto a priority: his denial of analytic status to "triangles have three sides" can be construed as a denial of a priori status to the corresponding proposition. Quine discusses analyticity, as opposed to a priority, because he is a hardline nominalist and empiricist who abjures propositions. In his view, talk of propositions is meaningless or, at best, unrigorous – only talk of sentences passes methodological muster. So we can meaningfully talk of analyticity (necessary sentences) but not a priority (necessary propositions), except by way of talking of analyticity.[363]

The problems with Quine's argument start right there. Quine obviously *does* think there are such things as sentences; so he does countenance the notion of a sentence. But we need the notion of *meaning*, of propositionality, if we are to help ourselves to the notion of a sentence. *Sentence* is a semantic category; things qualify as sentences by virtue of having a certain semantics – by virtue of having certain meanings. But once we have the notion of meaning , we have, I submit, the notion of truth *wholly* in virtue of meaning.

If two sentences have different meanings, then something has to *follow* from the one which doesn't follow from the other. Suppose that S and S' have different meanings but nothing follows from the one which doesn't

follow from the other. In that case, S and S' tell you exactly the same thing. If nothing follows from S that doesn't follow from S', and vice versa, then from what S tells you, you can conclude absolutely nothing that you cannot also conclude from what S' tells you, and vice versa. Now if you can conclude nothing from what S tells you that you cannot conclude from what S' tells you, and vice versa, then what S tells you coincides with what S' tells you. And in that case, S and S' give you exactly the same message; they have the same meaning. So if S and S' are to have different meanings, that entails – if doesn't actually coincide with – there being something that *follows* from the truth of the one sentence that doesn't follow from the truth of the other. But if you grant that something *follows* from the truth of a sentence, you *ipso facto* grant that that sentence *entails* something – that it is associated with analytic truth.

The obvious response to this is:

Nothing follows form '2+2=4' that doesn't follow from 'arithmetic is incomplete'; for they are analytically equivalent. But those sentences have different meanings. So there is some error in the logical structure of the argument you just gave.

Technically – but *only* technically – this is accurate. The basic idea behind our argument survives this attack. But our attack does need to be refined a bit. When two sentences have exactly the same analytic consequences, but different meanings, this is because – to put it very roughly – the *way* the one sentence relates to those consequences is different from the *way* the other relates to those same consequences. The logical derivation of '3+3=6' from '2+2=4' would be dramatically different from the logical derivation of '3+3=6' from 'arithmetic is incomplete.' The analytic route, so to speak,

linking ‘2+2=4’ and ‘3+3=6’ is very different from the analytic route linking the latter with ‘arithmetic is incomplete.’ What this tells us is that the meaning of a sentence is determined *not only* by what its analytic consequences. (Obviously the meaning of a sentence is a function of that *in part.* For if S and S’ have different analytic or logical consequences, then *ipso facto* they have different meanings.) The meaning of a sentence is determined *also* by the ‘analytic’ or deductive *routes* that link it to those consequences. (That, no doubt, is bound up with the fact that analytically equivalent propositions can have very different ‘cognitive values’ or ‘senses’; the cognitive value of ‘2+2=4’ is very different from that of ‘arithmetic is incomplete’.)

But the larger point is that there is no way to make sense of the notion of *meaning* except in terms of the notion of ‘following from’ – except in terms the notions of entailment and thus of analyticity. ‘Bob is bald’ has a different meaning from ‘grass is green’ because what follows from the one is different from what follows from the other. ‘Arithmetic is incomplete’ and ‘2+2=4’ have different meanings because, even though the same propositions follow from both, they follow *in different ways* – the analytic route between the one sentence and a given one of those consequences will be very different from the analytic route that consequence and the other sentence. So, again, even when we are dealing with analytically equivalent sentences, the notion of meaning is still to be understood in terms of that of entailment. So once the notion of meaning is countenanced, the notion of entailment (analyticity) is countenanced, and Quine’s project aborts. But if Quine is to have access to the category of a sentence – and it is hard to see how he can do without this category – then he needs the notion of meaning.[364]

There is another reason Quine cannot help himself to the notion of a sentence without committing himself to the existence of meanings. A

sentence obviously cannot be defined as a mere physical type (as a type of sound or type of inscription). Suppose that in language L the sound "snowiswhite" means *flying penguins play checkers poorly*. Further, suppose that the *syntax* of this L-sentence is totally different from that of the homophonous English sentence; so the L-sentence 'snowiswhite' doesn't parse into 'snow' and 'is' and 'white', and there are numerous inflections whose phonetic realization is totally different from that of their English counterparts, if they even have such counterparts. (So, for example, the 's' at the beginning shows that the sentence is in the indicative mood, the 't' sound indicates tense, and so on.) In that case, when someone speaking L produces the sound 'snowiswhite', he is surely not uttering the same sentence as an English speaker who says 'snow is white.' Surely sentences are individuated by, at the very least, how they parse syntactically.[365]. And once it is granted that sentences are individuated (in part) by their syntactic breakdowns, it becomes impossible not to allow that sentences are individuated by their *meanings*. For the syntactic structure of a sentence supervenes on its semantics. 'Snow is white' has the syntax it does because 'snow' is a noun, 'white' is an adjective, and so on. And obviously 'noun', 'adjective', and so on, are semantic categories. In other words, "snow is white" has the syntax it has because "snow" *picks out* a certain kind of substance and "white" *expresses* a property, and so on.

In sum, a sentence is a coupling of some physical type with a semantics. (There is bit more to it than that. Two very different sentences could bear the same proposition and involve the same physical type. For S and S' to be the *same* sentence, they must (i) involve the same physical type, and (ii) have the same *derivation* tree – basically, S and S' must be associated with the same proposition *in the same way*. But for our purposes it is enough to note that a sentence is *at the very least* a pairing of a physical type with a meaning:

obviously this truth is not invalidated by the fact that a certain *way* of effecting this pairing is necessarily involved.) And it isn't really clear how the semantics of a sentence could be anything other than a proposition.

Of course, different authors may have different views as to what it is for a sound-type to *bear* a certain proposition. Platonists will say that this consists simply in its being associated in a certain way with a certain non-spatiotemporal entity. Others will want to say that it consist in the sentences being *used* a certain way. (Arguably, these two views coincide: the thing which *associates* the meaning 'red' with *redness* is that the former is used in a certain way – that tokens of it occur in consequence people having certain mental contents under certain circumstances. So there is not, as far as I can tell, any conflict between the idea that 'meaning supervenes on use' and the idea that meaning is some platonic entity: the platonic entity is assigned to an expression *by way of* its being used a certain way. But, of course, I am on controversial ground here.) But there is no denying that sentences are individuated by (at least in part) the propositions they bear and, consequently, that 'sentence' is a semantic category.

3. Quine's entire argument, as we'll now see, essentially involves his construing sentences as merely *physical* types.

His argument is basically this. What is it for a sentence to be analytic? It is, Quine answers, for it be a logical truth ('all single males are males') *or* for it to be a truth that can be made into a logical truth by substituting synonyms for synonyms. (We can turn 'bachelors are males' into a logical truth by substituting 'single male' for 'bachelor'.) A 'logical truth' is, in this context, is one whose truth can be determined by considering syntax alone. (Quine's definition of analyticity is, I will argue later, is too narrow; I will argue that, although 'a circle is a closed planar figure of uniform curvature', it cannot be

turned into what Quine's calls a 'logical truth' by substituting synonyms for synonyms. But let us leave that aside for now.)

But, Quine argues, synonymy can only be defined in terms of analyticity. Two expressions are synonymous only if the one can be substituted for the other without changing the modal status of the sentence in question – without, in effect, turning an analytic truth into a synthetic one or *vice versa.* So 'rhenate' is not synonymous with 'chordate' because substituting replacing 'rhenate' in the sentence 'all rhenates have kidneys' with 'chordate' turns an analytic sentence into a synthetic one: 'all rhenates have kidneys' is analytic, and 'all chordates have kidneys' is synthetic. So analyticity must be defined in terms of synonymy, but synonymy must be defined in terms of analyticity. Here we have a vicious circle. So, Quine concludes, there is *no* definition of 'analyticity' – there is no intelligible category corresponding to that term. To characterize a sentence as 'analytic' is nonsense.

4. This brings us to the next leg of Quine's argument. Given that there is no such thing as analyticity, any sentence is revisable in light of experience. Quine infers from *this* putative fact that only the class of accepted sentences *as a whole* 'faces the tribunal of experience.' What is the basis for this inference?

When we decide to *reject* a given sentence in light of experience, that is because we are tacitly accepting certain other sentences. If I decide to *reject* some astronomical theory in light of an experience I have, that is because I am *accepting* various other sentences -- -sentences describing the behavior of light, sentences giving the physics of the telescope I using, and so on. So if we choose to reject a sentence S_1, that is always because we are accepting *other* sentences $S_2 \ldots S_n$. If I choose to reject the sentence 'there is an elephant in the room' even though I have a visual experience of an elephant,

that is because I *accept* other sentences, for example 'I am prone to hallucinations', 'recently I have been using psychedelic drugs', and so on.

Since an experience E compels us to *reject* a sentence S_1 only if we accept other sentences $S_2 \ldots S_n$, we can with equal rationality reject $[S_2, \ldots S_n]$ and hold onto S_1. In fact, given E, we could with equal rationality hold onto $[S_1, S_2, \ldots S_n]$ by rejecting some *further* set of sentences $[S_{n+1}, S_{n+2} \ldots]$. So any given sentence can be held onto; any given sentence can be made to be insensitive to further experience. In fact, any subset of the set of sentences can made invulnerable to experience by jettisoning some of the sentences not falling into that set.

But obviously which sentences we countenance *is* at some level sensitive to our experiences; we countenance 'dogs don't fly' because of what our sense-experience tells us. Since any given sentence, and any subset of the set of sentences, can be made insensitive to experience, we must conclude that it is the set of accepted sentences *as a whole* that is sensitive to new experiences.

Since any subset of the set of accepted sentences is only *conditionally* responsive to new experience – the condition being that sentences *outside* that set are held onto – we must suppose that experience bears unconditionally *only* on the class of accepted sentences *as a whole.* So what we believe *is* sensitive to observation. Further, given that there is no such thing as analyticity, what we believe is sensitive to nothing *other* than observation. (For surely a sentence that is made true by something other than observation is analytic or a priori.) But with regard to the sentences that embody these beliefs, it is only these sentences *taken as a whole* that is amenable to experiential confirmation or disconfirmation.

5. Evaluating Quine's argument

There is excellent reason to suspect that there is *some* error in Quine's argument. It is pretty clear that 'squares have four sides' is analytic, and that *squares have four sides* is a priori. And it would be easy to generate other propositions like this *ad infinitum*.

Kripke and Kaplan have shown the relationship between analyticity and a priority is not as straightforward as we used to think. But it seems pretty clear that *if* the proposition *all squares have four sides* is a priori, then the *sentence* "all squares have four sides" is analytic. Let us initially talk about a priority. Later on we will map what we have to say back onto analyticity. Right now I will show that some propositions *must* be a priori – e.g. *squares have four sides* – and that, consequently, some sentences are analytic (e.g. "squares have four sides").

It isn't really how concepts could be individuated *apart* from the a priori truths they subtend. How do you distinguish between two concepts C_1 and C_2 except by saying that what 'x falls under C_1' entails is different from what 'x falls under C_2' entails? If someone asks you 'what is the difference between a square and a rectangle?', you say 'if something is a square, that entails that it has four-sides of equal length, but if something is rectangle that doesn't entail that it has four sides of equal length.' So, by itself, the banal fact that concepts *differ* from one another seems to generate countless a priori truths.

Admittedly, Quine's nominalism gives him a certain invulnerability to this point: he simply denies that there are concepts.[366] But this very nominalism makes it very unclear what he is doing. Quine says that no 'sentence' is unrevisable. By 'sentence' Quine surely cannot mean 'physical token paired off with a *proposition* of some kind.' For if one is committed to propositions, then one is surely committed to concepts. (If you concede that the proposition *triangles have three sides* exists, then you concede that the

concepts composing it – *triangle, side*…-- exist.) And if Quine accepts that there are concepts, it is hard for him to evade the arguments just given.

So by 'sentence' Quine surely cannot mean anything like 'a physical type coupled with a proposition.' So what *does* he mean? It seems to me by a 'sentence', Quine can only mean a *physical* type. (Of course, given that Quine is a nominalist, he will probably try to reduce types to tokens in some way. But we can leave that aside.) But it is obvious, and of no importance, that in *that* sense of 'sentence', any sentence can be revised.

6. Quine asks: what is it for two expressions to be synonymous? The answer is obvious: they're synonymous if they have the same meaning or semantic value. So there is no problem defining synonymy at all, contrary to what Quine argues. 'There are three feet in a yard' is analytic because 'yard' and 'distance of three feet' have the same meaning, the same semantics. Of course, once it is granted that there is a non-circular definition of 'synonymy', then Quine's argument collapses.[367]

Admittedly, Quine explicitly descries the doctrine that there are such things as 'meanings.' But then he opens himself up to the criticism stated in the penultimate paragraph.

This is not to mention the points made earlier: surely a sentence is something that has syntactic structure. It isn't *just* a physical type. And once it is granted that sentences have syntax, it must be granted that they have semantics. For the *syntax* of 'Bob loves Mary' owes to the fact that 'Bob' is noun, and 'loves' is a two-place predicate, and so forth. In other words, the syntax of that sentence derives from the fact that 'Bob', 'loves' etc. have certain *semantic* properties. Basically, the syntactic structure of a sentence derives from its semantic structure. Syntax is a *semantic* notion.

As Dummett astutely points out, Quine *himself* gives a definition of

analyticity, one that doesn't make any use of the concept of synonymy: a sentence is 'analytic' if it is not revisable in the light of experience.[368] Of course, Quine denies that there are any such sentences. But his reason for that denial is that there is no (non-circular) definition of 'analytic'. So given that Quine himself gives such a definition, it isn't really clear what justification he has for rejecting as nonsensical the category of analyticity.)

7. Let us suppose, with Quine, that whenever experience induces us to reject a sentence, that is only because we accept other sentences.[369] (This is certainly true for most sentences, though arguably experience does seem to give a *direct* justification for sentences like 'I am in pain'. But we can leave this aside.) If this is correct, then any given sentence S_1 is justified *indirectly*: by way of other sentences S_2 …S_n. Now E, along with S_2…S_n justify S_1 by virtue of there being some kind of *probability-relation* holding between [E, S_2…S_n] and S_1: the former *confirms* the latter. In any case, confirmation relations – relations of probabilification – must surely hold among sentences. Otherwise, experience really would be inert in terms of justifying the acceptance or rejection of sentences. But are these relations of probabilification given by *empirical* propositions? On pain of a vicious regress, they cannot be.[370] Suppose it were the case that, *for any two propositions or pieces of information P_1 and P_2*, the bearing that P_1 had on P_2 were conditional on some further piece of information P_3, then the bearing of P_3 on P_1 and P_2 would itself be conditional on some further piece of information P_4 – so nothing would ever manage to probabilify anything.

This no doubt demands clarification, for it will at once be said:

Whether P_1 bears on P_2, and if so, to what extent, is a function what *besides* P_1 we accept. If there is evidence that all crows tend to attack humans, and I see a crow, then the proposition

(N) *there is a crow nearby*

will give high probability to the proposition

(A) *we will be attacked shortly.*

But if there is no such evidence, (N) *nearby* will not confirm (A). Of course, it is an empirical matter whether there is such evidence. So what probability *there is a crow nearby* gives to *we will be attacked shortly* is an empirical matter.'

This is quite wrongheaded. To say (N) does not *by itself* confirm or disconfirm (A) is to say that taken by itself (N) *objectively* has *no* bearing on (A). To say that, given evidence E, (N) confirms (A) to degree .8, is to say that the conjunction of (N) and certain other propositions $P_1 \ldots P_n$ jointly give probability .8 to (A), where $P_1 \ldots P_n$ are the propositions that give E.

Of course, *further* evidence could reveal that our danger of being attacked by the nearby crow was actually .001 – we might encounter some new evidence E* to the effect that this particular crow had never attacked anybody. But that doesn't mean there has been a change in the confirmation-relation between A on the one hand, and the conjunction of N and $P_1 \ldots P_n$, on the other hand. That confirmation relation hasn't changed at all. What it means is that, when are attempting to predict whether that crow will attack us, we should consider *not* the probability relation just described, but some

other one, where this other one is this: *the conjunction of N and $P_1 \ldots P_n$ and $Q_1 \ldots Q_n$ gives probability .001 to (A), where $Q_1 \ldots Q_n$ are the propositions that embody the information contained in E**. This, I think, disposes of the objector's point.

But we have not shown that the sentences/propositions giving probability relations are *not* revisable in light of experience; we have only disposed of a bad argument to the effect that they are thus revisable. Could probability relations – and we can include entailment relations among these – be revisable in light of experience? It is hard to see how. Whenever experience warrants the rejection or acceptance of a proposition, it does so *by way of* probability relations holding between that evidence (or, more accurately, among the propositions in which that evidence is embodied) and that proposition. So propositions affirming probability-relations must, ultimately, have a non-experiential – an a priori – justification.

For the sake of argument, suppose that any proposition/sentence of the form 'P_1 probabilifies P_2 to degree x' was empirical and was thus revisable in light of experience. That would mean that only *experience* could tell you what probability P_1 gave to P_n. So let's say you have some experience E_1 that is relevant to the bearing P_1 has on P_n. It might seem as though this settles matters: 'I have some experience E that tells me what probability P_1 gives to P_n. Where is the problem for Quine?' But this only pushes the problem back. Experience will give us only *another* set of propositions, and we will need *further* experience to know the probability relation between them. Let's say I have experience E, whose content is embodied in proposition P_2. On Quine's account, I now need to know the probability-relation between [P_1 and P_2], on the one hand, and P_n on the other. If Quine is right, then this relation is not knowable *a priori*; I need *experience* to tell me what it is. So let's say I have another experience E*, whose content is

embodied in some proposition P_3. Will this settle matters? No – because, if Quine is right, I need experience to know the probability relation between [P_1 and P_2 and P_3], on the one hand, and P_n, on the other. Of course, this continues *ad infinitum*

The point is that if probability relations between propositions (or, if you prefer, sentences) are always a posteriori, then nothing can ever succeed in confirming anything. Whether P_1 confirms P_2 will always be conditional on whether some *other* proposition is accepted. But then the probability relation between *that* proposition and the first two will be conditional on our accepting some *fourth* proposition – and then some fifth, and so on *ad infinitum.* Probability relations must be a priori – must be fixed independently of experience – *if* any experience is to bear on our beliefs system *by way of them.*

What we just said about probability relations can be mapped onto entailment relations. Actually, entailment relations are arguably limiting cases of such relations. (Hempel held this, but it is controversial.) Suppose that entailment relations – e.g. *if P, and (if P then Q), then Q* – was revisable in light of experience. In that case, of course, that proposition/sentence could only be validated by experience. But it is hard to see how experience could validate *any* belief – apart from those that follow directly from sense perception (assuming there even are any such beliefs) – except by way of modus ponens.

So Quine's holism, which seems to follow from his rejection of the analytic/synthetic distinction, makes no sense unless we accept that *there are* a priori (or, if we think in terms of sentences, analytic) truths. For it is only via such truths that sense-perceptions have any impact on propositions that are *not* on the 'periphery' of the 'web' that is our belief system.

8. This brings us to another point about Quine's holism. Quine thinks (i) that there are no analytic truths and (ii) that this implies that some kind of holism about our beliefs must be the case.

But really the opposite is the case. If there are no analytic truths, then a very extreme form of doxastic *atomism* must be the case.

This is because it is not just sentences (or propositions) affirming *entailments* that are analytic: it is also, as we've just seen, sentences affirming probability relations. Any relation of bearing is given by an analytic sentence: not just those where the antecedent gives 100% probability to the consequent. Probability relations are the medium through which observation sentences alter the body of our beliefs. So if we get rid of all analytic sentences/propositions, then a given observation will be absolutely inert. Far from affecting the totality of our beliefs, it would have *no* bearing on anything; it would not – *pace* Quine – have bearing on the totality of our beliefs. For, to sum up, bearing-relations are given by analytic, not synthetic, sentences. [371]

9 Bonjour has a crisp refutation of Quine, which is similar to one we gave a moment ago:

Could an argument of any sort be entirely justified on empirical grounds? It seems clear on reflection that the answer to his question is 'no.' Any purely empirical ingredient can, after all, always be formulated as an additional empirical premise. When all such premises have been explicitly formulated, either the intended conclusion will be explicitly included among them or it will not. In the former case, no argument or inference is necessary, while in the latter case, the needed inference clearly goes beyond what can be derived entirely from experience. Thus we see that the repudiation of all a priori

justification is apparently tantamount to the repudiation of argument or reasoning generally, thus amounting in effect to intellectual suicide. [372]

This is a compelling argument. Here is another way of putting it. Obviously not everything we know about the external world comes *directly* from sense-perception – sometimes, maybe even all of the time -- there is an element of inference. Even if you are merely *remembering* something you saw, the justification for your believing in the proposition in question has an inferential component. (The transition from *I remember P* to *P is the case* is, strictly speaking, inferential, as Hume pointed out.[373])

Let P_n be such a proposition. By hypothesis, P_n is *inferred* from what we sense-perceive (it is not like my belief that there is a table in front of me, which I *see* to be true). Let $P_1 \ldots P_{n-1}$ be the propositions embodying the sensory information on the basis of which we infer that P_n is true. What is the status of the proposition *[$P_1 \ldots P_{n-1}$] implies P_n*? Is it a priori or is it a posteriori? If that proposition is not empirical – i.e. if it is a priori (analytic) -- then there is a priori truth (and Quine is wrong). So let's assume that it is not. In that case, P_n must be a proposition that you can know to be true *directly* on the basis of what you see. But by hypothesis, it is *not* such a proposition: a contradiction. So to avoid contradiction, we must assume that there is a priori (or, if we think in terms of sentences, analytic) truth.

10. Quine doesn't deny that "all bald males are males" is true: for it is *formally* true, and he concedes that *formally* true sentences are analytic. But Quine thinks that he can concede the existence of analyticities of *this* type and *not* concede the existence of non-formal analyticities like "all cases of

knowledge always cases of belief".

But this shows a failure to understand the concept of "formal truth". Why is "all bald males are males" true? Because it encodes the proposition: *all bald males are males* and that proposition is true. That *proposition* is formally true – true in virtue of its form – and its having that property is merely *reflected* in the sentence "all bald males are males." But it is not the fact that the sentence has a certain *form* that makes it true, except in a rearguard sense: the sentence has a form that reflects the form of the *proposition* it bears, and that *proposition* is formally true.

"All bald males are males" is surely a "logically true sentence" if ever there was one. It encodes the proposition *all bald males are males.* That, and that alone, is why it is true. If "all bald males are males" meant *penguins have hands* – which it could – it wouldn't be true.

We could invent a language in which *all bald males are males* was not given by a sentence of the form "all things that are both F's and G's are F's". Suppose that we added a new word to English – "blurbos" – which, by our stipulation, is to have the same meaning as "male." In that case "all bald males are blurbos" would be true for the exact same reasons as "all bald males are males." But "all bald males are blurbos" would, Quine's reckoning, be in the same camp as "all knowledge is true belief": its truth would not be "formal', for it would not have the sentential form "All F's and G's are F's." So it is kind of thing that Quine would regard as a faux-analytic truth. But it is obviously is an analytic truth: for what makes it true is the exact same thing which makes "all males are males" true. By parity of reasoning "all knowledge is true belief" is analytic.

11. If you try to show that the analyticity of

‘the predecessor of 2 is identical with the unique number n such that, in a Euclidean space S, given a point P in S, and line L passing through P, then for any point P* not on L, exactly n lines pass through P* that don’t interest with L’

reduces to some synonymy, you will inevitably *presuppose* rules like (if P then Q, and if Q then R, then P then R). But presumably Quine cannot presuppose the validity of such analyticities, since his aim is to show that they don’t exist.[374]

The other problem is that this proposal in question (analyticity, if it exists, reduces to synonymy) is subject to obvious counter-examples. The sentence ‘if x is green, x is colored’ is presumably analytic. But it is well known that “green” is not defined by anything of the form “….colored…” You can, so to speak, add the property of being male to the property of being unmarried and get the property of bachelorhood. But you cannot *add* something to the property of being colored to get the property of being green.[375] So whatever it is that underwrites the truth of “if x is green, then x is colored”, it is not a synonymy relation between “green” and some expression of the form “…colored…” – it is not a synonymy relation at all.

This point generalizes. Consider the analytic truth “the measure of the sum of the interior angles of a hexagon is larger than the sum of the interior angles of a triangle.” Is the truth of this underwritten by a synonymy relation? The only way to *convert* that truth into one that holds in virtue of a synonymy relation is by doing a fair amount of mathematics, and thus by presupposing various analyticities. So analytic truth is *not* to be understood in terms of synonymy.

11. There is a related problem with Quine's point that analytic truth is always underwritten by *synonymies.* If that were really the case, then analytic sentences would not be informative or, if they were informative, they would not be informative about *concepts*[376] but only about language. But analytic truths are often deeply informative.

The truths "there are three feet in a yard" and "a fortnight is a period of two weeks" give you no *insight* into the concepts *yard* or *fortnight.* If they are informative, it is because they are *showing* you that the *word* "fortnight" is synonymous with "period of two weeks", and that "yard" is synonymous with "distance of three feet." But it is not *typically* the case analytic truths are only *verbally* informative. Surely "there are continuous curves that cannot be differentiated at any point" is not informative on a merely verbal level; it gives you an insight (a highly counter-intuitive one) into concepts. But if analytic truths held wholly in virtue of *synonymy* they would be informative *only* about words.

Another example may help. Consider the sentence:

(S) "x is a class of points equidistant from a given point in a three-dimensional space iff x is a closed three-dimensional figure of uniform curvature"

That sentence is analytic. But surely the expression

"class of points equidistant from a given point in a three-dimensional"

is not a mere notational variant of – is not *synonymous* with – the expression

"closed three-dimensional figure of uniform curvature".

Those two expressions are as *non*-synonymous as they could be. They have different Fregean senses, and have totally different decompositions. (One decomposes into the concept *point*; the other does not.) That is why S is informative. So Quine's basic assumption – synonymy underwrites analyticity – is just false.

12 Could Quine's argument – leaving aside its other shortcomings – be changed to accommodate the fact that analyticity doesn't typically reduce to synonymy? I don't see how. The notions of entailment and, therefore, of analyticity are *propositional* (or, if you prefer, sentential) notions: they apply to whole propositions (sentences). The essence of Quine's attack is that these notions cannot be non-circularly grounded in the right kind of sub-sentential notion, i.e. in terms of some notion that applies, not to sentences, but to sub-sentential entities (e.g. nouns). The only sub-sentential notion that could, with even remote plausibility, serve this argumentative function would indeed be that of synonymy. So once it is conceded that synonymy is not adequate to this task – the reason being that analyticities are not always, or even typically, underwritten by synonymies – it is hard to see how any argument even vaguely like Quine's would be up to the task.

13. In *the Philosophy of Logic*, Quine defined a 'logical truth' as one that comes out true under any reinterpretation of its non-logical constants.[377] So 'all unmarried males are males' is logically true because no matter what terms we substitute for 'male' and 'married', the result is a truth (so long as the substitutions are uniform). So a logical truth is one whose truth-value depends only on the meanings of the logical constants in it: everything else occurs only vacuously.

Sometimes this position is phrased as follows: logical truths are purely *formal* truths. But this way of putting it is misleading. The truth of ‘no unmarried males are not males’ obviously depends on the *meanings* of ‘not’, ‘no’, ‘are’. When it is said that a sentence is true ‘in virtue of its form alone’, that is really an obscure way of saying that the *semantic* facts which make it true or false can be read off the *syntax* of that sentence. So, in the end, it is the *semantics* of ‘no unmarried males are not males’ which makes it true.

Also, the *syntax* of that sentence is not separable from its semantics; to know the syntax of that sentence, one must know the *semantics* of ‘no’, ‘not’, ‘are’ and any other logical connectives occurring therein. Syntax is not morphology; I can know the *physical* shape of a symbol without knowing anything about its semantics; but I can know absolutely nothing about the *syntax* of a sentence without knowing something about its semantics – I must know the semantics of certain key terms. To know the syntax of ‘all non-glarfs are not not non-glarfs’, I must know the semantics of ‘all’, ‘non’ and ‘not’. Actually, I must even know *something* about the *semantics* of ‘glarf’; I must know that it is some kind of a referring expression. If I do not know this, then I have no assurance that ‘all non-glarfs are not not non-glarfs’ is a sentence at all and can thus make no deductions from it. If it turned out that “glarfs’ was an adverb, then the ‘sentence’ in question would be gibberish, and nothing could legitimately be deduced from it since, being nonsensical, it wouldn’t in fact entail anything. So I must know *something* about the *semantics* of “glarf” (and “blurb” and “Bob”…) if I am to know that “all glarfs are blurbs and Bob is a glarf” entails “Bob is a blurb”.

When we say that the ‘syntax’ of a sentence makes it true, that is really just an imprecise way of saying that a certain subset of the *semantic* facts characteristic of that sentence make it true. The identification of syntax with morphology – or with anything else that, like morphology, is totally non-

semantic -- is completely and totally bogus.

Let S be some sentence of Finnish that was homonymous with ‘that is a bird’, and was written out the same way. But suppose that S parses completely differently from ‘that is a bird.’ The ‘b’ is really a past-tense mark, the ‘th’-sound indicates the indicative mood, and so on. S would have an utterly different syntax from ‘that is a bird’ even though it had the same morphology. And the only way to know the syntax of S would be to have *some* knowledge of its semantics.

So the *semantics* of a logically true sentence are as much implicated in its being true as the semantics of ‘Bob smokes’ are implicated in *its* being true. But in that case, the distinction Quine wants to draw between ‘all bachelors are males’ and ‘all unmarried males are males’ vanishes. Both are true in virtue of semantics.

Undeniably, what makes ‘bachelors are male’ be true is its semantics. What makes “bachelors are male” be true is that its semantics, its meaning, is the proposition *all bachelors are male*, which of course is true. The same is true *mutatis mutandis* of ‘all males are males.’ Now Quine says that the *semantics* of sentences like ‘bachelors are male’ and ‘all cases of knowledge are cases of true belief’ doesn’t make them *necessarily* true. So on what grounds can he avoid saying the same thing about (e.g.) ‘all males are males’? He must somehow exempt the semantics of words like ‘not’, ‘all’, and so on, from his critique of analyticity. But on what grounds can he do this? His argument, if it works at all, works on ‘not’ and ‘all’ as much as it does on ‘knowledge’ and ‘bachelor.’

Consider the truth ‘not not P iff P’. We can map what Quine said about ‘bachelors are male’ onto this. ‘Not not P iff P’ is analytic because ‘not not P’ and ‘P’ are synonymous. But to say that they are synonymous is just to say that they are interchangeable *salva analycititate* – is just to say, in effect, that

statements like ‘P iff not not P’ are true. So there is no non-circular explication of what it is that makes ‘P iff not not P’ be synonymous.

I should make something clear. I *myself* think that ‘not not P’ and ‘P’ are *not* synonymous. ‘P iff not not P’ is indeed analytic. But, in my view, the analyticity characteristic of that sentence has nothing do with a synonymy relation holding between the terms flanking the ‘iff’. But *Quine* says that analyticity is grounded in synonymy. *Quine* says that ‘bachelors are unmarried’ is analytic only in so far as ‘bachelors’ is synonymous with some expression of the form of the form ‘…unmarried…’ So *Quine* would say that if ‘P iff not not P’ is analytic, that is because of a synonymy between ‘P’ and ‘not not P’. The argument given in the last paragraph shows that what Quine said about ‘all bachelors are unmarried’ applies equally to ‘P iff not not P’. So if Quine’s argument shows that ‘all bachelors are unmarried’ is synthetic, then that argument also shows that ‘P iff not not P’ is not analytic.

But ‘P iff not not P’ is surely analytic. Quine himself thinks so.

To this, Quine might say (in effect, he *does* say):

‘P iff not not P’ is *formally* true. ‘Bachelors are unmarried’ is not *formally* true. If the latter is analytic, that is in virtue of some fact about the *meanings* of the expressions composing it. More exactly, if it is analytic, that is because “bachelor” and some expression of the form “…unmarried…” are synonymous. The point is that, when we’re dealing with ‘all bachelors are unmarried’, we cannot ground analyticity in sentential *form*. So we need to ground it in a synonymy relation: and *that* kind of relation, as it turns out, must be grounded in an analyticity. So the non-analyticity of ‘all bachelors are unmarried’ has entirely to do with the fact that it is not formally true. But ‘P iff not not P’ is not *formally* true. So despite what you said, my argument doesn’t that *that* sentence is analytic.

But here I need only repeat what I said earlier. What does it mean to say that ‘P iff not not P’ is *formally* true? Obviously the ‘form’ in question is not *physical* form. That sentence could be realized by tokens that had practically any morphology at all. That sentence can be written in block letters, in script, in innumerable different fonts. It could be expressed in sign-language. There is no limit to how much the *physical* form of tokens of that sentence may vary. We could invent a code in which tokens of that sentence were realized by light-beams, sun-spots – anything at all. Clearly, then the ‘form’ Quine is talking about must not be physical form.

What else could it be? I can think of only one answer. It must be propositional form: the *form* of the proposition encoded in ‘P iff not not P’ is responsible for the truth of that proposition. So when Quine says that ‘P iff not not P’ is formally true, he must (if he is being coherent) be saying something about the proposition borne by that sentence. He must be saying that the *meaning* of that sentence has a certain property.

But once it is conceded that there are meanings, then Quine’s argument is destroyed at the outset. Once we countenance the notion of meaning, we can say: ‘bachelors are unmarried’ is analaytic because the *meaning* of ‘unmarried’ applies to anything to which the *meaning* of ‘bachelor’ applies. Quine says that the notion of meaning is meaningless: *that* is his justification for trying to ground analyticity in some kind of relation among *expressions*, as opposed to the things they mean; *that* is his justification for saying that, if ‘bachelors are unmarried’ is analytic, that is in virtue of a synonymy relation holding among its constituent expressions.

But we’ve seen that the notion of *formal* truth, no less than any other kind of truth, requires the notion of meaning. It is nonsense to try to reduce formal

truths to facts about physical morphology. When we say that “if P then P or Q” is a ‘formal truth’, that is a rearguard way of saying something about the *meaning* (proposition) that it bears.

So here is the position Quine is in. He can concede that there are formal truths, like “P iff not not P’ and “if P, then P or Q”. But in that case, as we’ve seen, he has to concede that there are propositions and meanings. And in that case, of course, his argument against analyticity is demolished. He cannot countenance meanings of any kind *while* holding onto the argument he gives in “Two Dogmas”.

The other option Quine has is this. He can hold onto his view that there are no propositions or meanings of any kind. But, to do that, he must *deny* that ‘if P, then P or Q’ and ‘P iff not not P” are formally true.

So, to hold onto the argument he gives in “Two Dogmas”, Quine must *deny* that that there are any formally true sentences. He must deny that ‘if P, then P or Q’ is formally true.

But if Quine denies that there are any formal truths, then what is he to say about the truths of logic? Quine cannot define logical truths as those that are formally true. And of course, he cannot define them as those that are analytic while *not* being formally true. What *can* he say about them?

Perhaps he might say that logical truths are just those which are very well entrenched. But such a position would involve our regarding as truths of logic things which, by our current conception, aren’t logical truths at all.

Let us sum up. If Quine’s argument is to work, then there are no formal truths. The position that there are no formal truths is obviously false. In any case, it has disastrous consequences.

14 Consider the proposition: *there are no analytic truths.* Is this analytic or

synthetic? Quine cannot say that it is analytic. So he must say that it is synthetic. But if he says that it is synthetic, then he is saying that its *negation* is not absurd: he is saying that *there are analytic truths* is not absurd. But Quine's whole point is that the proposition *there are analytic truths* is absurd.

The obvious rejoinder is this: 'Quine is arguing that the category of analyticity is bankrupt, not that it is uninstantiated, or even necessarily uninstantiated. He is not saying that 'there are analytic truths' is absurd, so much as he is saying that the proposition in question is a non-starter – that it involves an error so fundamental that neither it nor its negation says anything at all.'

But that is not very plausible. As Dummett says, Quine himself gives a definition of analyticity: *not revisable in light of experience.* Surely the italicized concept is a coherent one, even if it is not instantiated. (For *revisable in light of experience* is obviously a coherent concept – after all, Quine says that *every* sentence falls under that concept. So the complement of that concept – *not revisable in light of experience* – is also coherent.)

15 As we've seen, there is a very simple answer to the question: When are two expressions synonymous? They are synonymous iff they have the same meaning. Quine's argument gets going only because he rejects this answer, and he can reject this answer only because he rejects the notion of meaning. This raises the question: what, for Quine, *is* an expression? He cannot give the obvious answer: some kind of physical type (e.g. sound or inscription) paired with a meaning. For Quine, a language is cannot be a pairing of physical types with meanings. So what is it? Presumably he has a more naturalistic conception of what a language is: it is the totality of people's verbal behavior, or something of the sort. (By 'verbal behavior' I mean to include, not just overt bodily movements, but the mental activity that lies

behind them.) The platonic conception of language is: a language is a set of rules pairing meanings with expressions. The anti-platonic approach – which, presumably, Quine endorses – must see language as a wholly worldly phenomenon: people making certain noises, producing certain inscriptions, in consequence (presumably) of their being in certain psychological states. So for Quine (unless I am seriously misreading him), there is, ultimately, nothing to a language besides various bits of verbal behavior.

At first, this conception seems reasonable enough. But it breaks down as soon as we ask: what is *verbal* behavior? To behave *verbally* is not just to make noises with one's larynx or to produce inscriptions that have a certain shape. Laryngeal behavior is not necessarily *verbal* behavior. A brainless corpse could be electrically stimulated to produce the sound 'it's lovely out today'. But that sound-production would not be verbal behavior. What is missing?

At least *part* of what is missing is some kind of mental state. Presumably, for the production of a sound to qualify as *verbal* behavior, it must be caused by some kind of mental state.

But not just *any* mental state will do. Suppose we train a parrot to say 'it's hot in here' whenever it feels a slight increase in room temperature. Does that qualify as *verbal* behavior? Surely not. All we have there is stimulus and response. The parrot must *mean* something by that sound 'it's hot in here.' Given only that some mental state – a sensation of heat – *caused* that sound-producing behavior, it doesn't yet qualify as *verbal* behavior. When a dog howls in response to pain, it is not behaving in a *verbal* way, even though, no doubt, it's behavior is caused by a mental state (pain). (There is a difference between screaming in pain and *saying* "I am in pain." Saying "I am in pain" is a *verbal* act; screaming in pain is not.)[378]

The same point *mutatis mutandis* shows that producing an inscription with

a certain shape is not *ipso facto* a case of verbal behavior. We could train a dog to inscribe the shape 'I am in pain' in the soil with its paw whenever it felt pain. But that would not qualify as verbal behavior.

For the production of a sound (or inscription) to qualify as *verbal* behavior, it must embody awareness of some *norm.* [379] When *you* say 'it's hot in here', that is *verbal* behavior – as opposed to non-verbal behavior, like kicking a chair in anger -- because (at least in part) your making that noise embodies your knowledge of various *rules.* You know that sounds of the form 'x is hot' are correct iff x has the property denoted by 'hot'. Basically, your producing the sound 'it's hot in here' is *verbal* behavior because your producing that sound embodies knowledge of various rules. What kind of rules? *Semantic* rules – rules like 'x is hot' is true iff x has the property denoted by 'hot' -- and thus rules that assign *meaning* to sounds. Of course, the same goes for your producing the corresponding inscription or (if you know sign-language) the corresponding hand-movements.

To sum up, the concept of *verbal* behavior is logically posterior to that of a *semantic rule.* And the latter concept, surely, involves the concept of *meaning.* So the kind of meaning-nihilism that Quine needs for his argument to work makes it impossible for him to countenance the notion even of *verbal behavior* and thus of *anything* linguistic. So Quine's argument is self-undermining.

16. In his great paper *Truth by Convention,* Quine argued compellingly against the thesis that logical truths are grounded in linguistic conventions. His argument was simple. Convention can directly account for the truth of, at most, a finite number of logically true sentences. But there are infinitely many logically true sentences. So the ones that don't follow *directly* from conventions must follow from conventions *and* logic. Thus conventionalism

reduces logic to convention *plus* logic, and doesn't reduce logic at all.

But Quine's dismissal of the notion of meaning seems to commit him to some version of the linguistic theory of logical truth. There is an obvious answer to the question: Why is the sentence "either snow is white or snow is not white" *true?* The answer is: it encodes the proposition *either snow is white or snow is not white* and that proposition is true. But Quine rejects the category of propositions; so he cannot give that answer. He can't say that what makes "either snow is white or snow is not white" is some non-linguistic *empirical* fact. It would be absurd to place this sentence in the same category as "snow is white". And Quine himself concedes this. He concedes that *formally* true sentences are analytic. So what makes the sentence "snow is white or snow is not white" be true, given that what does this is (i) not a non-linguistic empirical fact and (ii) not the existence of a logically true proposition. The only answer left to Quine seems to be: some fact about language; what makes that sentence true is *language*.

But now we can use Quine's own argument against conventionalism against *him.* Convention is responsible for, at most, the existence of finitely many formally true sentences. The rest follow from convention only in an indirect sense: they follow from convention *plus* logic. So facts about language can be *directly* responsible for the existence of only finitely many formally true sentences. The infinitely many remaining formally true sentences cannot be made true by facts about language alone. Thus they must be made true by something non-linguistic. That non-linguistic something cannot be empirical; Quine admits that formally true sentences do not have empirical truth-makers. (It would be absurd to say that "either snow is white or snow is not white" is in the same category as "snow is white". And Quine himself admits that there is a categorical difference between these sentences.) So that non-linguistic something must consist of principles of logic *conceived*

as non-linguistic entities.(They must be conceived as *non*-linguistic entities on pain of vicious regressiveness.) But if a principle of a logic is not a sentence, then what could it be other than a *proposition?*

Basically, to avoid some kind of absurd conventionalism or linguistic theory of formal truth, Quine must concede that principles of logic are *not* identical with sentences and are thus identical with *propositions.* So he must let propositions into his ontology. Now if he lets propositions into his ontology, then it is hard to see how he could avoid letting concepts (in the non-psychological, platonic sense) into his ontology. (For, presumably, propositions are constructed out of concepts and, possibly, objects.) But, as we've seen, as soon as he does that his argument is doomed.

17. A Davidsonian objection

There is another reason to countenance the existence of a priori (analytic) truth. This reason is much more conditional; it presupposes the correctness of a plausible, but very debatable, thesis.

People can only be understood in terms of rational norms. (This is the debatable thesis.) And it is hard to see how any substance can be given to the concept of rationality if we reject the idea that there are analytic truths.

The view I'm describing derives from Donald Davidson's work[380]; so let us turn to his words.

> [R]elations among the [proposition] attitudes are essentially logical: the content of an attitude cannot be divorced from what it entails and what is entailed by it. This places a normative constraint on the correct attribution of attitudes: since an attitude is in part identified by its logical relations, the pattern of attitudes in an individual must exhibit a large degree of coherence.

This does not, of course, mean that people may not be irrational. But the possibility of irrationality depends on a background of rationality: to imagine a totally irrational animal is to imagine a animal without thoughts. [381]

Permit me to unpack and supplement this statement. People are described normatively: as rational or irrational, as moral or immoral. This is not to say that people are not described in causal terms: they are. But where mental-causation (the event of one mental thing's causing another) is concerned, the causal order seems to track a kind of *rational* order. Bill believes that he must eat an apple to prevent a dangerous drop in his potassium levels Bill also believes there is an apple in Mary's lunch-box. The conjunction of these beliefs causes Bill to form an intention to open Mary's lunch-box, steal the apple, and so on. Notice that a causal nexus can be discerned between the initial condition (Bill believes such and such…) and the resulting event (Bill opens Mary's lunch box…) only in so far as some *rational* connection can be seen between them.

Where mental causation is concerned, the causal order seems to reflect the rational order. Obviously there is some divergence between the two. But Davidson's point – and it is not an unreasonable point[382] -- is that there cannot be a *total* divergence. Suppose someone has some belief B, and he infers on the basis of B that Timmy has four apples. In that case, B cannot possibly be the belief that Timmy has 39 apples. What *makes* something be a belief that Timmy has 39 apples just *is* (inter alia) that it *will* lead to beliefs like *Timmy has more than 38 apples* and that it won't lead to beliefs like *Timmy has four apples.* What it is for something to be a belief that Timmy has 39 apples just *is* for it to lead to certain beliefs and not to lead to others. If

we impute *enough* irrationality to someone, then we sin against that principle.

So the concept of rationality is actually built into the concept of a propositional attitude. If B leads to the belief that Timmy has 39 apples, B cannot be the belief that Timmy has no apples or that Timmy has four apples, because such a transition would be *too* irrational.

This is not an epistemic point. Davidson is not saying that B's leading to the belief that Timmy has no apples is good *evidence* that B is not the belief that Timmy has 39 apples. Davidson is saying: it is very hard to see what in virtue of what B could qualify as a belief that Timmy has 39 apples *if not* that B leads to beliefs like *Timmy has more than four apples* and *it is not the case that Timmy has four apples.*

If Davidson is right, then there are two reasons to believe in analytic truth. First, if Davidson is right, it is analytic that people will be rational up to a point. Second, if Davidson is right, a certain minimum level of rationality characterizes mental causation. For what *makes* B be (say) a belief in some proposition just *is* that B is caused by beliefs that entail P and that B causes beliefs that P entails. Of course, if Davidson's functionalism is correct, it very much seems that we need to countenance entailments and, thus, analytic truth.

Of course, one could try to empiricize these supposed entailments:

I admit that if B leads so and so to belief that Timmy has *no* apples, then *ipso facto* B is not a belief that Timmy has 39 apples. But that doesn't mean we have to regard *as analytic* the proposition

(T_1) *if Timmy has 39 apples, then Timmy does not have* no *apples (i.e. he has some apples*).

(T_1) is simply very well entrenched. And because it is so well entrenched, it is the kind of inference someone would make. Therefore B is not to the effect that Timmy has 39 apples.

But this line of thought isn't really consistent with Davidson's point. If (*) is simply well-entrenched, then in principle one *could* make that inference without being irrational. And this is what Davidson denies.

One could counter-respond by saying:

Given the kinds of beliefs *we* have, it would be irrational to hold (*). But given a completely different set of background beliefs, it wouldn't be irrational. Quine's point is that (apart from formally true sentences) *any* proposition *could* be true. So we *could* imagine some set of data $D_1 \ldots D_n$ that would make (*) true. Therefore, we *could* imagine a set of background beliefs that would make it rational to hold (*). Those background beliefs would be beliefs in the existence of $D_1 \ldots D_n$.

But in that case we *are* accepting an analytic truth. We are treating as analytic the proposition:

(T_2) *Given $D_1 \ldots D_n$, the truth of T_1 follows.*

If for each i, we treat T_i as merely well-entrenched, then we obviously incur a vicious regress.[383]

References

Boghossian, Paul Artin. 1997. “Analyticity”. In *A Companion to the Philosophy of Language.* Bob Hale and Crispin Wright (eds.). Oxford: Blackwell.

Bonjour, Laurence. 1998. *In Defense of Pure Reason.* Cambridge: Cambridge University Press.

Brandom, Robert. 1998. *Making it Explicit.* Cambridge: Harvard University Press.

Davidson, Donald. 1994. “Donald Davidson”. In *A Companion to the Philosophy of Mind.* Samuel Guttenplan (ed.).Cambridge, MA: Blackwell.

Dummett, Michael. 1973. *Frege: Philosophy of Language.* Cambridge: Harvard University Press.
--1978. “The Significance of Quine’s Indeterminacy Thesis”. In *Truth and Other Enigmas.* Cambridge: Harvard University Press.

Fodor, Jerry. 1998. *Concepts.* Oxford: Oxford University Press.

Hale, Bob. 1997. “Modality”. In *A Companion to the Philosophy of Language.* Bob Hale and Crispin Wright (eds.). Oxford: Blackwell Publishers.

Kaplan, David. 1989 “Demonstratives”. In *Themes from Kaplan.* Joseph

Almog, John Perry, Howard Wettstein (eds.). Oxford: Oxford University Press.

Kripke, Saul. 1972 *Naming and Necessity*.. Cambridge: Harvard University Press.

McCawley, James D. 1981. *Everything that Linguists have always wanted to know about Logic*. Chicago: University of Chicago Press

Mates, Benson. 1952. “Synonymy.” In *Semantics and the Philosophy of Language*. Leonard Linsky (ed.). Chicago: University of Illinois Press.

Pap, Arthur. 1958. *Semantics and Necessary Truth.* New Haven: Yale University Press.

Quine, Willard van Orman. 1966. “Truth by Convention”. In *The Ways of Paradox and Other Essays*. Cambridge: Harvard University Press.
--1968. *Philosophy of Logic.* Cambridge: Harvard University Press.
--1990. “Two Dogmas of Empiricism”. In *The Philosophy of Language.* A.P. Martinich (ed.). Oxford: Oxford University Press.

Wittgenstein, Ludwig.1953. *Philosophical Investigations.* Oxford: Basil Blackwell.
--1974. *Philosophical Grammar* . Rush Rhees (ed.). Oxford: Basil Blackwell.

Nomic Relations in Relation to Hempel's Paradox

Abstract: A raven-type paradox is generated by the co-existence of two statements that are logically but not confirmationally equivalent. I argue that, in order for such a paradox to occur, the statements in question must read as having nomic force. Otherwise they end up being confirmationally equivalent. This is because "accidental" (nomically innocent) generalizations are not confirmed by their instances, so that two accidental generalizations are vacuously identical in respect of their confirmational properties. But, as I attempt to show, once we expose the nomic content of the two statements in question, they turn out *not* to be logically equivalent on any delineation of the concept of a nomic connection. So Hempel's paradox can be dealt with within the framework of classical logic, and without sacrificing our any of our intuitions regarding confirmation (in particular, our pre-theoretic intuition – denied by Hempel, among others -- that the existence of a red carpet fiber does *not* confirm "copper conducts electricity" or "all ravens are black").

I. (*) "all ravens are black"

is logically equivalent with

(**) "all non-black things are non-ravens."

Logical equivalence guarantees confirmational equivalence: if two hypotheses H1 and H2 are logically equivalent, then any given piece of

evidence confirms H1 exactly as much as it confirms H2. (**) is confirmed by pink trousers and white pianos. But (*) is *not* confirmed by such things. So we have two statements that are logically but not confirmationally equivalent. A paradox.[384]

Some philosophers bite the bullet and say that pink trousers and white pianos *do* confirm (*) as much as they confirm (**). Others deny this, and try to solve the paradox through the construction of non-standard logics – logics in which (*) and (**) are *not* logically equivalent and which therefore accommodate the confirmational differences between them. A number of solutions combining features of both these proposals have been proposed. These solutions are drastic. It is hard to believe that pink pianos provide positive support for (*). And it is *ad hoc* to invent a whole new system of logic to deal with this one puzzle. (That would be like an accountant creating a new system of arithmetic to accommodate a bookkeeping error.) So, understandably, none of the aforementioned solutions has met with universal acceptance. I would thus like to propose a different approach.

II. The sentence:

(AC) "all the coins in Smith's pocket are quarters"

is said to express an "accidental generalization". The sentence:

(*)"all ravens are black"

does *not* express an accidental generalization. (*) is to the effect that, if a thing is a raven, that provides some kind of nomic basis (though, obviously,

one that is extremely far from being sufficient) for its being black. So (*) affirms the existence of a nomic connection between two properties. We might say that it expresses a "nomic generalization".

There is an important difference between accidental and nomic generalizations. The latter do, while the former do not, receive confirmation from their instances. If I pull a quarter out of my pocket, that gives me no reason to believe that the next coin I pull out will be a quarter. The next coin I pull out could just as well be a dime or a nickel. So (AC) is not confirmed by its instances.[385]

But if I see a raven and it is black, that *does* give some reason to believe that the next raven I see will be black. Unlike (AC), (*) *is* confirmed by its instances.

Now consider the statement:

(CA) If a thing is not a quarter, it is not a coin in Smith's pocket.

Like (AC), (CA) expresses an accidental generalization, at least on its most natural reading. So, by reasoning exactly similar to that just given, non-quarters that are not coins in my pocket are no confirmation of (CA). (CA) is not confirmed by its instances.

(AC) and (CA) are logically equivalent. They are also confirmationally equivalent. (AC) is not confirmed by quarters drawn from my pocket. Nor is it confirmed by coins that are not quarters that are not in my pocket. (CA) is not confirmed by coins that are not quarters that are not in my pocket. Nor is it confirmed by quarters drawn from my pocket. There is nothing that confirms the one hypothesis more than the other, since there is nothing that confirms either.

Consequently, no raven-type paradox arises in connection (CA). Because

(CA) and (AC) lack nomic force, neither is confirmed by its instances. That prevents them from being confirmationally different. That in turn prevents the occurrence of a raven-type paradox. For such a paradox to occur, we need two statements that are logically equivalent and confirmationally *non*-equivalent. (CA) and (AC), at least when they are read as devoid of nomic content, are logically *and* confirmationally equivalent.

Before moving on, we should deal with a possible worry:

Imagine the following scenario. Every time my friend Smith pulls a coin out of his pocket, it is a quarter. This happens unfailingly for twenty years: I have never seen him pull a coin out of his pocket that was *not* a quarter. On the basis of this information, I surely have some reason to believe that the next coin he pulls out of his pocket will be a quarter. So surely the coin-withdrawals that I witness *do* provide for the statement that there are no coins other than quarters in Smith's pockets. So I don't see why (AC) is different from (*) in this respect.

In the case described, the statement "there are never any coins other than quarters in Smith's pockets" is *not* an accidental generalization, and it *does* have nomic content. In the case in question, what you come to believe is precisely that there is some kind of principled connection between being a coin in Smith's pocket, on the one hand, and being a quarter, on the other. You come to believe that Smith's policy is never to have coins other than quarters in his pockets (because, perhaps, when he has to feed the meter or do laundry, he doesn't want to deal with the frustration of rummaging through useless nickels and dimes), or that he is the victim of a long-running practical joke. In any case, what you come to believe is *precisely* that the situation in question is not happenstance. Of course, it is not a *law of nature* that any coin

in Smith's pocket must be a quarter. But that doesn't mean that no nomic connection is at work. It is not a law of nature that, whenever you flip the switch, the light turns on. But there is obviously a nomic connection at work. Flipping the switch *causes* the light to turn on: the first event provides the nomic basis for the second. There is a principled, nomic relation between the first event and the second. Similarly, if Smith's mission in life is to make sure that there are never any coins in his pocket that are not quarters, or if Smith's roommate wants nothing more than to steal all coins other than quarters from Smith's pocket, then there will be a causal mechanism ensuring that the property of being a coin in Smith's pocket is co-instantiated with the property of being a quarter. There will be a nomic, principled relation at work.

Like any other "all"-statement,

(AC) "all the coins in Smith's pocket are quarters"

is capable of two readings. On one of them, it is to the effect that there is some kind of *connection* between being a coin in Smith's pocket, on the one hand, and being a quarter, on the other.

There is another possible reading of (AC). On this reading, it is to the effect that every coin in Smith's pocket just happens to be a quarter, the concomitance being purely "accidental".

On the *first* reading, (AC) clearly receives support from its instances. The proposition that there is a *principled* relationship between the property of being a coin in Smith's pocket and the property of being a quarter is one that *is* supported by each instance of a coin in Smith's pocket that is a quarter.

On the second reading, (AC) is *not* capable of such confirmation. On that reading, (AC) has no nomic content, and makes a purely numerical statement. If there really is no principled connection between the property of being a

coin in Smith's pocket and the property of being a quarter, then *ex hypothesi* each case of a quarter being withdrawn from Smith's pocket must be considered to be entirely independent of each other such case. In that case, tautologously, we cannot use those past cases as a basis for a prediction regarding future cases. In thinking otherwise, one would be guilty of the so-called "gambler's fallacy."

Past data is confirmationally inert except in so far as it suggests the operation of mechanisms that can be counted on to continue. So past data can confirm a hypothesis only if that hypothesis has nomic content – only if it is to the effect that the correlation in question has a basis in law. Statements like (AC) are thus capable of confirmation only when they are given a nomic reading. Really this is a truism. To give (AC) a nomic reading just *is* to take it to say that the concomitance in question is rooted in the kind of mechanism that warrants our making predictions on the basis of past information. To give (AC) a non-nomic reading is to read it as *not* saying that the concomitance in question has any such basis. So a non-nomic reading is, practically by definition, one that doesn't warrant prediction and, therefore, doesn't allow confirmation. To sum up, it turns out to be a truism that genuinely accidental generalizations do not receive confirmation from their instances.[386]

If (*) and (**) both lacked nomic force, then they would both be incapable of being confirmed by their instances. Consequently, they would be equally incapable of confirmation by their instances. In that case, they could not be confirmationally different, and there would be no paradox. There is a paradox only because at least one of those two statements has nomic force – only because at least one of them affirms the existence of a nomic connection between two properties. Of course, if one of them has nomic force, both must have nomic force if they are to be logically equivalent. After all, if P affirms a nomic connection and Q does not, then P and Q are not true in the same

models, and are therefore not logically equivalent. So if (**) is to be logically equivalent with (*), then we must see it as making a nomic statement.

So read in the way that is relevant to our inquiry, (*) is to the effect that:

($*_N$) If a thing is a raven, that provides a nomic basis for its being black.

So if (**) is to be logically equivalent with (*), then we must see (**) as being to the effect that:

($**_N$) If a thing is non-black, that provides a nomic basis for its being a non-raven.

The question now is: Are ($*_N$) and ($**_N$) logically equivalent? If not, then there is no paradox: for the paradox in question to arise, we need ($*_N$) and ($**_N$) to be confirmationally non-equivalent while being logically equivalent.

I will now argue that, in fact, ($*_N$) and ($**_N$) are *not* logically equivalent. For reasons previously discussed, their non-equivalence prevents the paradox in question from being generated.

III. There are different kinds of nomic relations. The most obvious examples of nomic relations are *causal* relations. Consider the statement:

(#) Anything[387] that drinks arsenic at time t is dead by time t*.

This is a true statement (at least for certain values of t and t*), and it is

obviously not an accidental generalization. There is a nomic connection between drinking arsenic at time t and being dead by time t*. This nomic connection is causal: drinking arsenic *causes* death. So the nomic content implicit in (#) is given by the statement:

($\#_C$) Drinking arsenic at t causes death at t*.

Some nomic relations that seem *not* to be causal actually turn out to be causal. Consider the proposition:

(PF) Pinewood floats.

(PF) is not an accidental generalization. It is, in fact, a natural law, albeit a very derivative one. The law in question is given by a *causal* statement:

(PFC) If something is pinewood, that is causally sufficient for its floating when placed in water.

Let $P_1 \ldots P_n$ be the properties individuative of pinewood. It is pretty clear that, where at least some of these properties are concerned, a thing's having those properties prevents it from sinking when placed in water. In other words, a thing's having those properties causes it to float. So the law embodied in (PF) is actually a causal law, even though (PF) does not initially appear to be a causal statement.

At least arguably, not all nomic relations are causal. Consider the statement:

(WO) "all water has oxygen in it".

(WO) is not an “accidental generalization”. It expresses a principled relationship. But that relation is not causal. Being composed of water doesn’t *cause* a thing to be composed of oxygen, and being composed of oxygen doesn’t *cause* a thing to be composed of water. The relation in question is not one of causation, but of *constitution.* For something to be water is (inter alia) for its constituent molecules to be composed of oxygen atoms.[388]

IV. It is time to take stock. Let O be an arbitrary object, and let phi and psi be arbitrary properties. We’ve considered two ways that O’s having phi might provide a nomic basis for its having psi. There may be a *causal* relation between O’s having phi and its having psi. Or there may be a relation of *constitution.* If a thing has the property of drinking large quantities of arsenic at time t, that *causes* it to have the property of being dead at t*. If a thing has the micro-structure characteristic of pinewood, that *causes* it to float when placed in water. If something is made of water, that does not *cause* it to be composed of oxygen atoms: its being composed of water *consists* in its having such a composition.

Are there nomic connections that don’t involve relations of causation or of constitution? Some new terminology will help us answer this question. Drinking arsenic at time t is one thing. Dying at time t* is another. Since they happen at separate times, and possibly separate places, the one event is not in any sense an ingredient of the other. So we might say that a thing’s drinking arsenic does not *consist* in its dying, and that its dying does not *consist* in its drinking arsenic.

Let us now answer the question posed a moment ago. Suppose that a thing’s having phi does not consist, even in part, in its having psi. So a thing’s having phi is one thing, and its having psi is separate. They happen at

different times or different places or both. We are thus dealing with distinct states of affairs – states of affairs such that neither is a component, proper or improper, of the other.

Given this, suppose that there is some nomic connection requiring a thing to have psi if it has psi. That connection must plainly be *causal.* Given any two distinct states of affairs, any law requiring the existence of the one, given the existence of the other, is *ipso facto* causal. A causal relation just *is* something that requires the existence of one state of affairs, given the existence of some other state of affairs. So it is a truism, or at least follows from truisms, that under the circumstance described, a thing's having phi must be *causally* related to its having psi. [389]

Now suppose that there is no causal relation between a thing's having phi and its having psi, but that there nonetheless *is* some kind of nomic connection: a connection that demands the existence of the one, given the existence of the other. If a thing's having phi were *distinct* from its having psi, then for reasons we've seen, any nomic connection of the kind described would be *causal.* So if there is to be a non-causal, nomic connection between the two, that must involve a thing's having phi *not* being distinct from its having psi – the one state of affairs must be a constituent of the other.

It seems to be a truism, or at least to follow from truisms, that any nomic relations between a thing's having one property and its having another must either be *causal* or must lie in the fact that possession of the one property is *constitutive* (at least in part) of possession of the other. Nomic relations are relations of causation or of constitution.

V. Now we can begin to close the argument. We saw that each of (*) and (**) must be given a nomic reading if there is to be a raven-paradox. On such a reading, (*) is to the effect that there is some kind nomic relation between a

thing's being a raven and its being black. As we've just seen, such a relation would be one either of causality or of constitution.

Suppose that the relation is one causality. In that case, the nomic relation in question would be given by a story like the following. The class of ravens forms a natural kind. Let $R_1 \ldots R_n$ be the properties individuative of this kind. Where some subset of these properties is concerned, if a thing has the properties in this subset, that causes it to grow black feathers. If a thing has certain genotypic properties distinctive of ravens, that is a cause of its growing black feathers (at some early point in its life – perhaps before it hatches).

Now suppose that the nomic relation between a thing's being a raven and its being black *is* not causal. In that case, the nomic relation in question would be given by a story like the following. Supposing that $R_1 \ldots R_n$ are the properties individuative of this kind, there is some i such that R_i *is* the property of being black.[390]

We saw that, if there is to be a paradox, then (**) must be given a nomic reading – otherwise (*) and (**) will not be logically equivalent. There are two possibilities. The nomic connection can one of causality or of constitution. If the connection is one of causality, then (**) is to the effect that there is a *causal* relation between being non-black and non a non-raven. If the connection is one of constitution, then (**) is to the effect that for something to be a non-raven just *is* (inter alia) for it to be black, just for something to composed of water *is* (inter alia) for it to be composed of oxygen.

Given these points, it is clear why, on the relevant readings, (*) and (**) cannot be logically equivalent. Suppose we give causal readings to (*) and (**). In that case, (*) would be to the effect that:

($*_C$) Being a raven (having the genotype characteristic of ravens) *causes* a thing to be black (to grow black feathers).

And (**) would be to the effect that:

($**_C$) Being non-black *causes* a thing to be a non-raven.

($*_C$) and ($**_C$) are *not* logically equivalent. There are logically possible worlds, or models, in which a thing's having the genotype distinctive of ravens *causes* it to have black feathers, but where being non-black doesn't *cause* things to be non-ravens. An illustration may be in order. In the actual world, ($**_C$) is false. My shirt is non-black (it is green). But its being green did not *cause* it to be a non-raven; it never was, and was never going to be, a raven; nothing had to prevent it from becoming a raven, or make it stop being a non-raven. By contrast, ($*_C$) may well be true; in any case, it is an empirical question whether it is true or not. And given only that ($**_C$) is false, it cannot be *deduced* that ($*_C$); given the falsity of ($**_C$), it remains an open empirical question whether ($*_C$) is true or not. So ($*_C$) and ($**_C$) are not logically equivalent.

So, read causally, (*) and (**) are not logically equivalent, and there is no raven-paradox. What about if we give them the other possible reading? In that case, (*) would be to the effect that:

($*_K$) A thing's being a raven *consists in* (among other things) its being black (just as a thing's being composed of water consist in, among other things, its being composed of oxygen atoms).

And (**) would be to the effect that:

(**$_K$) A thing's being non-black *consists in* (among other things) its being a non-raven.

There is an obvious difference between (*$_K$) and (**$_K$). (*$_K$) is meaningful. In fact, it is true. But (**$_K$) seems to be false. Intuitively, it seems false to say that that a thing's being non-black *consists* in its not being a raven. Of course, it may be that all non-black things *are* non-ravens. It may even be that this is necessarily the case. But it still seems wrong to say that being non-black *consists in* being a non-raven.

Let us try to be more explicit about the logic underlying these intuitions. One wants to say that being non-black is not *about* being a non-raven: rather, it is *about* having a surface that doesn't absorb certain wavelengths of light; or perhaps it is about looking a certain way in certain lighting conditions. If you want to say what it is for something to be black, you don't mention the fact that it is not a raven; from an explanatory viewpoint, being non-black is not "about" being a non-raven. By contrast, if you want to say what water is, you *do* mention the fact that it contains oxygen. Being water is "about" (inter alia) being composed of oxygen atoms; if you want to say what it is for something to be water, you talk about oxygen atoms (and hydrogen atoms).

Evidently, the relation of "consisting in", in the relevant sense, is one that holds between x and y when there is an explanatory relation between them. (*$_K$) says (among other things) that there is an explanatorily significant connection between the property of being water, on the one hand, and the property of being composed of oxygen, on the other. The words "consist in" mark the presence of an explanatorily significant relation. The alleged

relation does exist, and that is why ($*_K$) is true.

The categories *non-raven* and *non-black* are not explanatorily significant ones, at least not in remotely the same sense as the categories *raven* and *black*. Given only the information that something is *not* a raven, or that it is *not* black, our best theories don't really tell you anything about it. They leave it open whether it is a blue diamond, a green iguana, a point in space, or a prime number. That information puts no predictively significant constraints on what that thing might be. But given only that that something is a raven, or that it is black, our theories have an enormous amount to say – about its bone structure, its physiology, its microstructure, its reproductive mechanisms, and so on.

The categories *non-raven* and *non-black* are explanatorily bankrupt. *x is a non-raven* isn't explanatorily significant. Neither *x is non-black.* But *x is a raven* and *x is black* are replete with explanatory content.

For this reason, ($*_K$) and ($**_K$) are not logically equivalent. For ($**_K$) to be true, the categories *non-raven* and *non-black* would have to be explanatorily significant. There are obviously possible worlds (our world is one of them) in which the categories *raven* and *black* are explanatorily significant, but where the categories *non-raven* and *non-black* are not. This means that there are possible worlds where ($*_K$) is true and ($**_K$) is not. So ($*_K$) and ($**_K$) are not true in exactly the same models and are therefore not logically equivalent.

There (infinitely many) other paradoxes of the type just discussed. But exactly similar considerations dissolve them.

VI. I would like to end by considering a possible objection to our analysis:

Your analysis involves the claim that all nomic relations are either

causal or constitutive. This is a very strong claim – one whose truth might well be doubted and therefore cannot be presupposed. And your argument is therefore shaky, at best, so far as it depends on it.

My argument is easily shown not, ultimately, to depend on the truth of the assumption in question (even though, to be sure, my use of that assumption did facilitate the exposition of my argument). My argument depends only on the assumption that statements such as

(PF) *pinewood floats* and (*) *all ravens are black* – non-accidental universal generalizations, in other words -- have nomic content of some kind or other.

We have argued that, when it is taken to express a natural law (or, in any case, some kind of principled, as opposed to merely accidental, generalization), (PF) has for at least part of its content the proposition:

(PF1) something's being pinewood provides some kind of nomic basis for its being something that floats

In light of this, consider the statement:

(PF2) If something doesn't float, then it isn't pinewood.

If (PF2) is read as having no nomic content, then it clearly isn't logically equivalent with (PF1) or therefore with the relevant reading of (PF). If (PF2) *is* read as saying that:

(PF3) Something's not being something that floats provides some kind of

nomic basis for its not being pinewood,

then (PF2) is *not* logically equivalent with (PF) (on the relevant disambiguation of the latter) and, moreover, is a highly dubious statement. (For reasons given earlier, it is hard to see how x's *not* being something that float could provide a *nomic* basis for x's *not* being pinewood.) It is true that, as we presented it, our argument involves the assumption that all causal relations are either causal or constitutive; and it may also true that this is a strong and doubtful claim. But all we need for our argument to go through is the claim that statements like (PF) and (*) (and *metal expands when heated* and *copper conducts electricity.*…) have some kind of nomic content. This is not an unreason able claim: it is not unreasonable to hold that *metal expands when heated* involve some kind of commitment to the position that a thing's being metal provides some kind of nomic, or at least principled, basis for its expanding when heated. Further, it is uncontroversial that, if *metal expands when heated* is read in this way, it is not logically equivalent with either *a thing's not expanding when heated provides some kind of nomic basis for its not being metal* or with any statement that *lacks* nomic content. Once this last point is granted, it straightforwardly follows that, read in the relevant way, *metal expands when heated* is not logically equivalent with *things that don't expand when heated are non-metals.* The same thing *mutatis mutandis* would seem to hold for any other pair of statements whose co-existence generates a raven-type paradox (e.g. *for all x, if x is pinewood, then x floats*/*for all x, if x doesn't float, then x isn't pinewood*). And therewith the Raven-paradox vanishes.

References

Ayer, Alfred Jules. (1972) *Probability and Evidence.* New York: Columbia University Press.
Cohen, Jonathan (1989) *The Philosophy of Induction and Probability*, Clarendon Press: Oxford.
Hempel, Carl. G. (1945a) *Studies in Logic and Confirmation.* Mind 54, 1-26.
Hempel, Carl. G. (1945b) *Studies in Logic and Confirmation. II.* Mind 54, 97-121.
Hempel, Carl G. (1966) *Philosophy of Natural Science.* Englewood Cliffs: Prentice-Hall. Falletta, Nicholas. (1983) *The Paradoxicon: a Collection of Contradictory Challenges, Problematical Puzzles, and Impossible Illustrations*.
Gardner, Martin (1973) *Mathematical Puzzles and Diversions*, Penguin Books.
Kuipers, Theo. (2000). *From Instrumentalism to Constructive Realism. On Some Relations between Confirmation, Empirical Progress, and Truth Approximation.* Dordrecht, the Netherlands: Kluwer.
Leavitt, F. (1996). "Resolving Hempel's Raven Paradox." Philosophical Inquiry 18, nos 3/4: 116.
Goodman, Nelson. (1954). *Fact, Fiction, and Forecast.* Cambridge: Harvard University Press.
Mackie, J.L. (1977). *The Cement of the Universe.* Oxford: Clarendon Press.
Maher, P. (1999). "Inductive Logic and the Raven Paradox". *Philosophy of Science* 66: 50-70.
Quine, Willard van Orman. (1969). "Natural Kinds". In *Ontological Relativity and Other Essays.* New York: Columbia University Press.
Sainsbury R. M. (1997) *Paradoxes*, (2nd Ed.) Cambridge University Press.
Scheffler, Israel (1981) *The Anatomy of Inquiry.* Indianapolis: Hackett.
Trusted, Jennifer (1979) *The Logic of Scientific Inference*, The MacMillan

Press Ltd.

A cure for Nathan Salmonella: How to do semantics without butchering the laws of logic

This paper will concern the following five contentions. (The first four are due Saul Kripke. The fifth is due to Nathan Salmon.)

(C1[391]) Names aren't definite descriptions (or encryptions or synonyms thereof).

(C2) The truth Hesperus is Phosphorous is both necessary and *a posteriori.*

(C3) Supposing there to be some one proposition P such P, and P alone, is what is meant by each two distinct sentences, somebody who understands both of those sentences can, at a given time, *rationally* accept the one sentence and reject the other. (So if P is what is meant by each of "Smith is evil" and "Jones is evil," one can rationally accept "Smith is evil" and rationally reject "Jones is evil."

(C4) One can, at a given time, both accept and reject some one proposition *and can do so rationally*. It is possible for there to be a proposition such that (for reasons having nothing to do repression or psychological fragmentation or any such condition) a given perso9n rationally accepts P and, at that very moment, rationally rejects it

C1 is true. In any case, we'll assume that it's true, if only for argument's sake. We'll find that the data usually thought to be warrant C1's rejection are entirely consonant with it.)

C2 is false. What about the more general claim that at least some

propositions are necessary and *a posteriori*? I won't attempt to refute (or establish) that claim. Suffice it so say that what we say in connection with C2 will suggest, though it may definitively prove, that C3 is false.

C4 is correct. A summary of Kripke's own argument for it, which cannot be improved on, will make that clear.

C5 is false. But what we'll see in the process of exposing the fallacies in Salmon's argument for C5 will be much more illuminating that the totally obvious, and therefore totally unilluminating, fact that C5 is false.

C3) Even if C2 is false, the class of necessary and *a posteriori* is non-empty. (It can known that the class of transcendental numbers is non-empty without anyone's being able to identify a single such number.)

C3 is probably false; but

C Kripke's first claim is correct. Names are not definite descriptions (or synonyms or encryptions thereof). Kripke's third claim is correct. Under some circumstances, one can understand two sentences, one of which affirms the negation of the proposition affirmed each of which one has a perfectly understanding accept a sentence that one understands and also accept

But what Kripke says about "Hesperus is Phorphorous" is only partly correct. The proposition meant by that sentence *is* necessarily true. So he's right about that. And the proposition that is typically *associated with* that sentence is non-trivial. (That non-trivial proposition identical, or equivalent with, one to the effect that:

(CO[392]) there is exactly one celestial object that is last to leave the morning sky and exactly one celestial object that is first to appear in the evening sky, and nothing that is a last celestial object to leave the morning skyh isn't also a first celestrial object to appear ikn the evening sky.) there is exactly one celestrial object appear first in the So *were* that *a posteriori*

proposition the one literally meant by “Hesperus is Phosphorous”--were it identical with the proposition that a moment ago we descreibed as “necessary”-- important astronomical discovery that couldn’t have been learned except through observation. contrary to what Kripke argues, the truth that Hesperus is Phosphorous is not *a posteriori*. The truth that is *conveyed* by utterances of “Hesperus is Phosphorous” is indeed nece

The first claim, we will find, is unqualifiedly false: simultaneously accepting a given proposition *and* its negation is the height of irrationality. When one takes into account some basic facts about the psychological antecedents of understanding, or intelligently using, a sentence, Kripke’s (1980) discovery (that names aren’t (encrypted) definite descriptions (or synonyms thereof)

To understand Kripke’s influential non-semantic claims, we must define four terms and also assimilate an important principle that Kripke discovered.

2.1 A priori vs. a posteriori knowledge The first two terms are ‘a priori’ and ‘a posteriori.’ These expressions are adjectives, and they modify expressions denoting knowledge. A priori’ knowledge is knowledge that one either has, or can acquire, merely by virtue of how one’s mind is structured; it can be thought of as knowledge that is coded into one’s cognitive structure or that follows from such knowledge. When you a buy a new computer, it typically comes ‘bundled’ with new software. A priori knowledge can be compared to such software.

2.1.2 Rigid vs. non-rigid designators The remaining two terms are ‘rigid designator’ and ‘non-rigid designator.’ For reasons that will emerge in due course, the distinction between rigid and non-rigid designators is meaningless

unless definite descriptions are taken to be referring terms. So in this context, we'll suppose, just as Kripke does in *Naming and Necessity*, that definite descriptions are singular terms.

Consider a world W that is semantically just like ours except that, in it, Al Gore, not Barack Obama, is the U.S. President in 2009. In W, 'the U.S. President in 2009' would refer to Al Gore. So *even though that expression would have the semantics that it has in our world*, it wouldn't pick out the same thing that it picks out in our world. In fact, given that, in W, Barack Obama isn't the president in 2009, if 'the U.S. President in 2009' d*id* pick out 'Barack Obama' in W, it couldn't possibly have the same semantics that it has in our world.

But suppose that, in world W* that is distinct from both W and our world, 'Barack Obama' referred to Joseph Biden. In that case, W* would *ipso facto* not be semantically identical with our world. In our world, there is some individual x such that x is Barack Obama and, as a matter of semantics, 'Barack Obama' refers to x. And any language in which 'Barack Obama' picked somebody else would to that extent semantically diverge from the specific version English actually spoken (even if it was otherwise just like it). Given that 'Barack Obama' refers to Barack Obama, it couldn't possibly *fail* to refer to him in a world that is semantically like ours. It could fail to refer to him only in a world where expressions were not in all cases governed by the semantic rules that govern them in our world.

'Barack Obama' is *rigid designator*. E 'rigidly designates' x if there is no possible world that is semantically just like ours where E fails to designate x. Proper names are rigid designators. Definite descriptions are not. 'The inventor of bifocals' refers to Bob Denver in some possible worlds, to Walt Disney in others, and so on.

Sometimes *demonstrative* expressions, e.g. 'that person', are described

as ‘rigid designators.’ The reasons for this are explained in Chapter__ Suffice it to say that they are not rigid designators. In fact, they are about as non-rigid as designators can be. ‘The inventor of bifocals’ is non-rigid in the sense that there are possible universes in which it refers to people other than Benjamin Franklin. But ‘that person’ refers to many different people in *our* world.

2.3 <u>Identity and necessity</u> Suppose that x and y are contingently identical – that, in other words, they’re identical but they don’t have to be. By Leibniz’s law, x and y have the very same properties. Obviously x is necessarily identical with x. Therefore, y has that property as well and is necessarily identical with x.

In general, identities hold necessarily if they hold at all. It isn’t a contingent fact that you are identical with you. If you exist, you can’t fail to be you.

Some sentences seem to be counterexamples to this, e.g.: the person standing over there is the person who owns the ice-cream store down the street.’ But that isn’t really an identity claim. We know from Russell that it’s real meaning is: something is both a person standing in that corner, unaccompanied by other people, and is also a sole proprietor of the ice-cream store down the street.’ Given any apparent exception to the principle that identity holds necessarily, if at all, reparsing that sentence shows it not to be an identity-claim.

A sentence of the form ‘A is identical with B’ expresses an identity claim *only* if ‘A’ and ‘B’ are rigid designators. Thus,

(IHP) ‘Hesperus is identical with Phosphorous’

is an identity claim. And so is:

(WH) ‘water is H2O’

and so is:

(MT) ‘Mark Twain is Samuel Clemens.’

2.4 Why Kripke believes that there are necessary a posteriori truths
Kripke made it clear that ‘Hesperus’ isn’t synonymous with ‘the morning star’ or with any other non-rigid designator, and he showed the same thing *mutatis mutandis* to be true of ‘Phosphorous.’ Given that Hesperus *is* Phosphorous, Hesperus can no more fail to Phosphorous than it can fail to be Hesperus. So given that it holds, IHP holds necessarily.

Kripke says that ‘IHP’ encodes an *a posteriori* truth. The idea is that it can’t be known *a priori* that IHP (or any other synonymous sentence) is true. It can be known only through astronomical observation. Given that IHP is necessarily true, it follows that IHP is both *necessary* and *a posteriori.*

2.5 <u>Evaluating Kripke’s attempt to show that there are necessary a posteriori truths</u> This is not a good argument. We made it clear earlier in this chapter that, although it doesn’t have such a claim for its *literal meaning*, IHP *communicates* a proposition of the form *the phi is identical with the psi.* In other words, it (non-literally) communicates a proposition of the form: *something uniquely has phi, and anything that has phi also uniquely has psi.*

So IHP communicates a *contingent* truth. It communicates, as we saw, the proposition that is expressed by the sentence:

(MS) ‘the last celestial body to disappear from the morning sky is identical

with the first to appear in the evening sky' (which we've abbreviated as: 'the morning star is identical with the evening star').

The proposition meant by MS is:

(MSP) something is uniquely a last body to disappear from the morning sky, and anything having that property also uniquely has the property of being a first thing to appear in the evening sky.

MSP is *a posteriori*. But it's also contingent. It wouldn't be true if matter were differently distributed.

Kripke is quite right that IHP has a necessarily true proposition for its meaning. But that proposition is also analytic and knowable *a priori.*

Here's why. We've agreed with Kripke that, for any property phi, 'Hesperus' isn't synonymous with ⌜the phi.⌝ And we've also agreed with Kripke, and with commonsense, that 'Hesperus' is a referring expression – that, in other words, there is some object x such that 'Hesperus' picks x out. Trivially, either 'Hesperus' picks out this object through the mediation of a description or it doesn't. In other words, supposing that O is what 'Hesperus' picks out, it either is, or is not, by virtue of O's fitting some description that 'Hesperus' picks out it. If it is, then 'Hesperus' is *ipso facto* a non-rigid designator – a definite description. For supposing that the description that O must satisfy is *unique thing having phi*, 'Hesperus' must be synonymous with ⌜the phi.⌝ But we know, from Kripke himself, that 'Hesperus' is not synonymous with any such expression. So 'Hesperus' does *not* refer to O through the mediation of a description. It picks O out – but *not* because it has this or that property. It picks it out because the semantic rule for 'Hesperus' is to the effect that, no matter what properties O has, 'Hesperus' is to pick it out

– that, in other words, ‘Hesperus’ picks out O, and this is contingent on O’s satisfying any condition (other than its being O). Thus, there is some object x (namely, O) such that the semantic rule for ‘Hesperus’ is to the effect that, no matter how O is, ‘Hesperus’ picks it out. In other words, ‘Hesperus’ *picks out O directly*. It is *directly referential.*

A consequence is, for some object x (such that x is Hesperus) such that, in virtue of having the form, ⌜Hesperus has psi⌝, a sentence is true if, and only if, x has phi. For exactly similar reasons, ‘Phosphorous’ is directly referential. It corefers with ‘Hesperus.’ Thus, there is some object x (such that x is Hesperus and Phosphorous) such that, in virtue of having the form, ⌜Hesperus has psi⌝, a sentence is true if, and only if, x has phi and such that , in virtue of having the form, ⌜Phosphorous has psi⌝, a sentence is true if, and only if, x has phi. Thus, ⌜Hesperus has psi⌝ and ⌜Phosphorous has psi⌝ *have the very same literal meaning.* In general, intersubstituting ‘Hesperus’ and ‘Phosphorous’ doesn’t’ change literal meaning. (This is entirely consistent with the fact such intersubstitutions may dramatically change suggested meaning, the reason being that one may learn what those two words refer to under very different circumstances and this in connection with very different bodies of descriptive data. The way the ‘referent is fixed’ will obviously affect that sentences containing the word in question connote, even though it won’t affect the literal meanings of such sentences.) Given that such intersubstitutions preserve meaning, it follows that

(IHP) ‘Hesperus is identical with Phosphorous’

has the same literal meaning as

(IHH) ‘Hesperus is identical with Hesperus’

(even though, given how it is typically made known what these words refer to, they have very different non-literal connotations). We’ve agreed that, for some x, each of ⌜Hesperus has psi⌝, and ⌜Phosphorous has psi⌝ is true just in case x has phi. It follows that, for some x, each of IHP and IHH is true just in case x has the property of being identical with x. Thus, for some x, IHP means:

(XSI) x is x.

And XSI is an analytic, a priori knowable proposition if ever there was one.

There is no a posteriori necessary truth. It appears otherwise only if one fails to distinguish sentences from the propositions that their meanings. The gulf between the two can be large, and metaphysics can’t be read off of sentential structure. This is the core idea of analytic philosophy. And it is a bit of shock that one of its most capable exponents has been remiss in applying, especially in such an important context.

3.0 <u>A cure for Salmonella</u>: Kripke (1979) wrote an article in which he put forth a new argument, one not stated in *Naming and Necessity*, as to why the Russell-Frege analysis of the non-triviality of IHP fails. Pierre is a monolingual speaker of French. He reads many books about London (which he knows as “Londres”). On the basis of the information contained in those books, he draws the rational inference that London is a beautiful city. Expressing this belief, he says:

(1) ‘Londres est jolie.’

It is to be emphasized that, given the data at his disposal, it is *rational* of Pierre to accept (1).

Our stories continues. Pierre is kidnapped and forced to live in a terrible part of London. He doesn't know that he's living in the city known to him as 'Londres.' He learns English, but he learns it as one learns a first language. In other words, he doesn't learn what English words mean by learning what their French translations are – he learns them the way that a native speaker of English learns them. He learns what 'snow' means, not by learning that it corefers with 'neige', but by learning a certain crystalline substance that falls on rooftops in the winter is called 'snow.' And he learns what "London' refers to in the same way. Given the data at his disposal, it is *rational* of Pierre believe that the sentence

(2) 'London is not pretty',

expresses a truth – a truth to the effect that the city he is living is *not* pretty. And Pierre does assent to it, and he utters that very sentence, knowing what it means.

Of course, (2) is the (English translation of the) *negation* of (1).

'London' is the translation of 'Londres.' So if 'London' is synonymous with some definite description -- with, let us suppose, 'the most populous city on the Thames' -- then 'Londres' is synonymous with the French translation of that definite description. But if *that* were the case, then (1) would mean the same thing as:

(1R) ' 'the most populous city on the Thames is pretty'

and (2) would be mean the same thing as:

(2R) ‘the most populous city on the Thames is *not* pretty.’

Pierre clearly *isn’t* in the same category as somebody who accepts *both* (1R) and (2R). So even if, as Russell and Frege thought, ‘London’ was synonymous with seom definite description, that wouldn’t explain the fact that (1) and (2) communicate such different propositions. In other words, it would fail to account for the *cognitive values* of those sentences. (The ‘cognitive value’ of a sentence is what it communicates.) Since the whole point of the Frege-Russell theory that proper names are(abbreviated) definite descriptions is to account for the cognitive values of sentences containing proper names, that theory fails to do what it’s supposed to.

3.1.1 <u>A cure for Salmonella (continued)</u> But Kripke’s parable exposes a problem much more profound than anything having to do with the semantics of a proper names. It’s a datum that Pierre understands each of (1) and (2), that he assents to each and sincerely utters each. So it’s a datum, or follows from data, that Pierre is *rationally* accepting some sentence *and* its negation.

This seems like a problem. A paradox. What’s the solution? Pierre is rational. No denying it. And Pierre understands each of (1) and (2). No denying that either. (To be sure, he doesn’t know as much *about* the thing that is referred to as ‘London’/’Londres’ as some other people. But so what? If I know who it is that ‘Larry’ refers to, then I know the *semantics* of ‘Larry.’ Obviously I don’t know who Larry is as a person as well as his therapist. But I know the *semantics* of ‘Larry’ every bit as well.)

Kripke says that he doesn’t know how to solve the problem. He says only that he’s inclined to suspect that. And ‘the apparatus of propositions breaks

down here.' He doesn't say what he means by that. (Nor does it matter, since, whatever it means, it's false.)

Nathan Salmon says that, given Kripke's parable, it follows that, at a given time, one *can* rationally accept a proposition *and* its negation.[393] Pierre would be rational, given what we've said so far, to believe that 'Londres^{1}London.' Thus, there is some x (namely, London) such that, according to Salmon, Pierre can rationally believe that: x^{1}x – i.e. that x is not self-identical. Thus, Salmon's position is that one can rationally believe that London is not self-identical – that, in fact, any given object is not identical with itself.

3.2. <u>Why, contrary to what Salmon ways, it is never rational to believe *P and not-P*</u> Salmon's solution is not only desperate, but incoherent. What is rationality? It is the ability to draw the consequences of the information at one's disposal. No truth is a consequence of its own negation, and every truth is incompatible with its own negation. Given that *P and not-P* is always false, it can never be rational to accept it or, therefore, to infer it. Salmon's position is inconsistent with the very meaning of the word 'rational.'

Salmon sometimes defends his position by saying that one can rationally accept *P and not-P* only if one doesn't know that what one is accepting a proposition *and* its negation. But this makes no sense. Your are irrational if you believe *P and not-P* or if you grasp both P and not-P and yet fail to generate the additional belief yet you fail to that they're inconsistent with each other. For acceptance of *P and not-P* to be rational, it would be necessary that the data at your disposal not to make it clear that you were accepting some proposition and its negation, which would be possible only if you didn't know what it was you were accepting. And this seems to be what Salmon holds.

But, contrary to what Salmon evidently thinks, you can't accept a proposition without knowing what it is. The counterexamples to this are spurious. For example, suppose that Smith sincerely says: 'Jones is supremely intelligent; so I believe whatever Jones believes.' If Jones believes that Lima is the capital Peru, does it follow that Smith also believes it? No. What follows is that *if* Smith were to know that Jones believed it, *then* Smith would believe it. In order to accept a proposition, it is not enough that it be *described* to you. You must actually grasp it. Another story will make this clear.

Person X, let us suppose, is never convinced of anything unless it actually is true. This has been verified, and I know this. Smith and I have a mutual friend, Green, who has been missing for the last three weeks. One day Smith tells me that he knows where our friend Green has been for the last three weeks. I have no idea where Green has been. I couldn't even guess. But Smith doesn't tell me where Green has been – only that he knows where he's been. So there is some proposition P such that, if correct, P says where Green has been and such that Smith accepts P. I know that whatever it is that Smith believes about Green's whereabouts is correct. But do I, under these circumstances, accept the proposition that identifies his whereabouts? No. If I were told what that proposition was, I *would* accept it. But I do not now do so. I don't know what the right value of 'P' is.

Suppose that, as Smith knows, Green has been in Spain for the last three weeks (but I'm not told this). Under these circumstances I don't accept the proposition: Green has been in Spain for the last three weeks. I *would* accept it if I knew that Green accepted it. But I don't yet know this, and I don't yet accept it. The proposition that I *do* accept is *this: I know that, whatever Smith believes to be the proposition describing Green's whereabouts over the last three weeks, that proposition must be correct.* But *that* proposition is very

different from *Green has been in Spain for the last three weeks.* And I don't yet accept the latter.

What this shows is that one cannot accept a proposition without knowing that one is doing so. When it *seems* as though one does this, one accepts some distinct proposition that describes some unknown proposition. If you accept a proposition, you know it. So if, for some P, you accept both P and not-P, you know it. But if you know it, you're not rational, since anyone who makes a faulty inference is *ipso facto* irrational (to that extent), and any inference whose conclusion is *P and not-P* is *ipso facto* faulty.

3.2.1 <u>Why, contrary to what Salmon ways, it is never rational to believe *P and not-P*</u> (continued) Salmon was driven to accept this, I dare say, preposterous view by his failure to take into account even the most basic points concerning the relationship between thought and language. Pierre doesn't learn what 'London' refers to in a vacuum. He is, we may suppose, air-dropped into a horrible part of a certain city. There is garbage all around. His captors are rude to him, and so are the people with whom his captors associate. (The only Londoners with whom Pierre's captors consort are those who, uncharacteristically for Londoners, are not very nice people.) He is then told: 'Welcome to London! This is your new home.' Let us suppose, in fact, that one of his captors is a philosopher of language who, like all those in that field, has a penchant for making things explicit. So, when first introducing Pierre to his new home, he says: "there is a certain city x, such that x is the city you are now occupying – this being the city whose dirty, garbage ridden streets you are now beholding – and such that 'London' refers to x."

Thus, as Pierre is beholding an awful, squalid, garbage infested quarter of London, his captor is says: 'London' refers to *this.* In being told this, what Pierre is being told is something along the lines of:

(L1) there is some vile squalid, garbage infested city x such that I am now in x and such that I was dragged here against my will and such that, because of x's obvious horridness, I am now depressed; moreover, 'London' refers to x.

An inevitable consequence of the fact that these are the circumstances under which Pierre learns what "London' refers to is that what an utterance of

(2) 'London is not pretty',

would impart (not semantically encode) is something along the lines of:

(L2) there is some vile squalid, garbage infested city x such that I am now in x [etc.] and such that 'London' refers to x; moreover, x is not pretty,

and, in general, that for any property psi

(L*P) ⌜ London has psi⌝

would impart (not semantically encode) is something along the lines of:

(L*P2) there is some vile squalid, garbage infested city x such that I am now in x [etc.] and such that 'London' refers to x; moreover, x has psi.

We've already made it clear why, even though L*P would *convey* L*P2, Pierre wouldn't be under any illusions as to what the former literally meant. It's also obvious that, given that since it is through L*P Pierre knows what 'London' refers to, it is rational of him to accept (2).

Let's think back to the time, before he was kidnapped, when Pierre was still in Paris. As we've said, Pierre has a lot of information about London (about the city he knows as 'Londres') and he draws the only reasonably inference that, given that information, he can draw, namely that it is a pretty city. He has read in various books that he knows to be authoritative that that:

(##) There is a certain city x such that x is more populous than any other in Britain, such that is the Capital of Britain, such that there are beautiful buildings in x, many wonderful and aesthetically pleasing cultural landmarks (e.g. Big Ben, the Royal Palace, Westminster Abbey), such that x's streets are generally very clean, there isn't much violent crime in x, there are beautiful buildings in x, etc.; moreover, 'Londres' refers to x.

Given that it is through some proposition along the lines of ## that Pierre knows what 'Londres' refers to, it is inevitable that an utterance of

(1) 'Londres est jolie'

would convey to him something along the lines of:

(##^) There is a certain city x such that x is more populous than any other in Britain [etc.] and 'Londres' refers to x; moreover, x is pretty.

And given that (1) communicates (##^) to Pierre, it's clear, given the information at his disposal, that he's rational to accept (1).

Here's what's going on. Pierre does know what 'London' refers to and also what 'Londres' refer to. He doesn't know that that they refer to the same thing, since the body of descriptive information that he uses to attach

‘London’ to the right thing is different from the body of descriptive information that he uses to attach ‘Londres’ to that thing. But he still knows what ‘Londres’ and ‘London’ refer to. (I may not know that my friend Smith is also the masked bandit; but I know who ‘Smith’ refers to.) And, given that Pierre speaks both French and English, he understands (1) and (2). He rationally assents to both, even though the one is the negation of the other. This is *not* because he rationally accepts some proposition *and* its negation. *Nor* is it because ‘the apparatus of propositions breaks down’, whatever that means. No – it’s because, given the information *through* which he learned the relevant semantic rules (the ones that, for some x (namely London), are to the effect that ‘Londres’ refers to x and that ‘London’ refers to x), (1) and (2) communicate very different propositions to Pierre. Those propositions don’t entail each other, and they are, in addition, supported by very different bodies of data.

There is no proposition that Pierre both accepts and rejects. Salmon is just plain wrong to think otherwise. And his contention that one *can* rationally accept propositions and their negations involves a number of rather obvious fallacies. First, sentences are not propositions. Suppose that sentence S1 encodes proposition P1 and that sentence S2 encodes the negation of P1. Given *only* that information, it doesn’t follow that a person who understands S1 and S2 can’t rationally accept both of those sentences. For it could be that, given the information through which that person learned the relevant semantic rules, the *propositions* conveyed to him by those sentences *could* both be true. Salmon agrees that, given only that the proposition encoded in the one sentence is the negation nof the proposition encoded in the other, one can rationally accept both sentences. But *he* thinks this is because one can rationally accept *propositions* that are inconsistent with each other; Salmon thinks that one accept the *propositions* P *and* not-P *without being irrational.*

Which is itself a deeply irrational belief. Irrationality just *is* acceptance of incoherencies, such as P and not-P.

Salmon holds, correctly, that Pierre is rationally in accepting both (1) and (2). According to Salmon, it follows from this that Pierre is rationally accepting some proposition *and* its negation. But that would follow *only* if what (1) and (2) conveyed to Pierre were confined to what they literally mean. But what sentences convey isn't always confined to what they literally mean. It's seldom confined to it. Every semanticist knows this. It's unclear why Salmon doesn't know it or, if he does, why he didn't see its bearing on this issue.

No Such thing as a Meaningless Sentence: Erin McCready Refuted

1.0 "Colorless green ideas sleep furiously" absurd, but meaningful In his review of my (2007) book *Conceptual Atomism and the Computational Theory of Mind*, Erin McReady says that

(MS) "colorless green ideas sleep furiously"[394]

is a *bona fide* sentence that is meaningless. Apart from citing MS, McCready does little to substantiate his claim that there are meaningless *bona fide* sentences.[395]

MS is absurd, not meaningless. MS can be taken to mean that there is nothing that is a colorless green dream that isn't *also* something that sleeps furiously. Thus read, MS is true (albeit vacuously so) and therefore meaningful). MS can also be taken to mean that there are colorless green dreams and, moreover, that any such thing sleeps furiously. Thus read, it's

false for a number of reasons (e.g. any green thing *ipso facto* isn't colorless, dreams don't sleep, things that sleep don't do so furiously). Another possible reading is that a thing's being a colorless green dream disposes it for some reason or other (e.g. for reasons of mathematical or natural law) to be something that sleeps furiously. (As they are ordinarily understood, "triangles have three sides" and "metal expands when heated" are, respectively, to the effect that, for reasons of logic (or mathematics or whatnot) a thing can't be a triangle without having three sides and that, for reasons of natural law, a thing can't be metal if it doesn't expand when heated." [396])

2.0: Absurdity vs. meaninglessness

"Aargdsahsdflkjh" is meaningless. It isn't absurd. In response to an utterance of it, one can't say: "that's absurd!" It cannot appear as the penultimate line of a legitimate indirect proof. By contrast,

(SC) "there are square circles"

is absurd, and one *can* describe an utterance of it as preposterous; it *can* occur as the penultimate line of an indirect proof; and, importantly, it can be embedded in true statements (e.g. "it's not the case that there are square circles," "if *per impossible* there were square circles, there would be objects that both had, and lacked, uniform curvatures"). SC is absurd *because* it's meaningful. It's absurd by virtue of the fact that there is a proposition P such that P is SC's meaning and such that P is inherently incapable of being true. [397]

Russell (1940)[398] claimed that:

(QP) “quadruplicity drinks procrastination”

is a *bona fide* sentence that is meaningless. (Quadruplicity is the property being a four-membered set; or so Russell says---rightly, we’ll assume.) Russell’s supposedly meaningless sentence is false, since the property of having four members, being non-spatiotemporal, cannot drink anything. Being false, it’s meaningful.[399]

Though probably driven in part by a failure distinguish between ridiculous meanings and non-meanings, Russell’s belief that QP is meaningless was probably also motivated by his acceptance of the Theory of Types, which, in Sections 4.0-5.2, we’ll discuss and, in agreement with contemporary logicians, find to be false and, indeed, absurd.

. According to Ryle (1949),

(OU) "Oxford University is in place p,"

is meaningless, since Oxford University cannot, according to Ryle, have a location (the reason being, Ryle says, is that universities are individuated by their functional, not their geographical characteristics).[400]

Ryle’s reasoning defeats itself.. Ryle says that OU doesn’t have a location. So he says that, for any place p, OU doesn’t p. So, by Ryle’s own olights, OU is false, not meaningless.

In claiming that

(TS) “This stone is thinking of Vienna”

is literally meaningless, Carnap (1932) is making the now familiar mistake of confusing ludicrousness with meaninglessness. Stones don't think. So TS if false. (And if they do think, TS is true.)

CG, QP, OU, and TS are all meaningful. So far as they can non-hyperbolically be described as "nonsense," it's because "nonsense" has ceased to mean "without sense," and has come to mean "having an idiotic sense," or some such.

2.1 The embedded clause test Each of these sentences can meaningfully occur as a proper part of larger, indisputably meaningful sentences, for example:

(MSR) "if colorless green ideas sleep furiously, it follows that there are things that are both colorless and green."

MSR is a garden-variety counterlogical statement, like "if there is some pair of numbers p and q such that p/q have no common divisor and such that $(p/q)^2=2$, then p and q both have, and don't have, a common divisor." If the occurrence in MSR of CG were meaningless, MSR would be meaningless. Since MSR is meaningful, so is CG. Exactly similar points hold of QP, OU, and TS.

Some sentences are brazenly absurd. Others are absurd but not obviously so. We use obviously absurd ones, like CG, to root out the absurd ones that *seem* reasonable, thereby preventing the latter from infiltrating and rotting out entire departments of thought.

2.2 Non-words that mimic words In many lectures of his that I attended

(as a UCLA undergrad), David Kaplan, a well known semanticist, insisted that the opening sentence of Lewis Carroll's Jabberwocky (" Twas brillig, and the slithy toves/Did gyre and gimble in the wabe") was a *bona fide* sentence. It is not. It is a sentence-impostor. (Obviously "brillig," "slithy," "tove," etc. are meant to bring to mind the words "brilliant," "slimy," "toad," etc. But those two lines no more token a sentence than a man wearing a Richard Nixon mask is Richard Nixon.)

In general, there are expression-like entities that, given their phonetic properties along with facts about the linguistic environments in which they occur, are easily mistaken for bona fide expressions. Given what the words flanking it are and what it sounds like, the occurrence of "blerko" in "sometimes a blerko can be a man's best friend" is likely to be mistaken for an actual word. But it has no meaning and is thus no word.

2.3 Words that, in certain contexts, aren't functioning *as* words and therefore, in said contexts, *aren't* words Oftentimes, a homonym of an actual word occurs in a context in which no semantic rule assigns any meaning to it and in which it therefore means nothing, even though the context in question is one that is otherwise just like a sentence. What occurs in "the absolute is one" is not the word "absolute," but a mere homonym of that word. For the occurrence of "absolute" in "the absolute is one" to be a word, there would have to be some semantic rule that assigned it a meaning. Although there are semantic rules that assign meaning(s) to homonyms of that occurrence, that occurrence per se is assigned no meaning. Anything that is assigned no meaning is ipso facto not a linguistic expression. So "the absolute is one" is not a sentence, at least not until definite values are assigned to "absolute" and "one." Until then, the only difference between it and ‹the phi is psi›[401] is that, whereas "phi" and "psi" are used *exclusively*

as variables, “absolute” and “one” are not used exclusively as variables. (A related point is that, when “absolute” and “one” are functioning as variables, their ranges would seem to be circumscribed by the meanings they have when functioning as constants.)

2.4 An important meta-semantic corollary Tautologously, anything to which no semantic rule assigns a meaning *ipso facto* isn’t a linguistic expression. I am contradicting myself if tell you that “blerb” is a word of English and then proceed to tell you that there is no semantic rule of English that assigns meaning to it (or its tokens). Obviously “blerb” can be flanked by the same sort of verbiage as actual expressions-----cf. “Jerry is a real blerb”---and, as this very example shows, might, depending on what the verbiage is, be naturally taken to have this or that meaning (in this case, one to the effect that Jerry has some sort of negative attribute, I should think). But until a semantic rule assigns it meaning, it’s not an expression; and once it becomes meaningful---once, say, “blerb” becomes synonymous with “creep” (which could happen: cf. “nerd,” “dweeb”)---a semantic rule (viz. for any x, *‹x is a blerb› is true iff x is a creep*) *ipso facto* assigns it a meaning.

The semantic rules of English do not assign a meaning or referent to the token of “absolute” in “the absolute is one.” So in that context, “absolute” isn’t functioning as a word and is, it would seem, functioning as, if anything, a (restricted) variable (a variable whose substituends are presumably restricted to expressions denoting things that are august and important in some way).

Israel Sheffler[402] tries to explain why:

(GW[403][404]) “Some vessels contain glubbified water” [405]

isn't "scientifically" or "empirically" meaningful, whereas

(HW) "Some vessels contain boiling water" [406]

is, by pretty much any standard, meaningful.

The only solutions to this problem that Scheffler acknowledges are ones which take it for granted, first, that (GW) is a *bona fide* sentence, albeit a meaningless one and, second, that it's because the occurrence of "glubbified" in GW renders that sentence unfit for verification or falsification or confirmation.

The actual solution is a lot more obvious than Scheffler's discussion suggests. "Glubbified" means nothing. There is no semantic rule of English that assigns a meaning to it. It's not that it's assigned an unscientific or unempirical meaning. It's simply assigned no meaning at all.

Suppose there were some rule R of the English language to the effect that for x to be "glubbified" is for x to contain little demons, whose existence couldn't be verified or falsified or confirmed, that, when present in a body of water, rendered that water unfit for human consumption. In that case, GW would be meaningful. Useful to science? Probably not. Meaningful? Yes. Why? Because, if it weren't meaningful, we couldn't judge that, *given its meaning*, it was empirically unverifiable(/unfalsifiable/etc.); we couldn't judge it to be unscientific or otherwise unmeritorious.[407]

3.0 Which sentences are meaningful?[408] This question is easier to answer than it is has been made out to be. Here is the answer: S is a meaningful sentence if S says that x has phi, for some object x and property

phi, and is thus equivalent to a sentence of the form: ‹x has phi.›

Any given sentences is either (i) non-molecular (e.g. “Smith is tall,” “Smith is afraid of Jones,” or (ii) a quantified generalization (e.g. “some people smoke,” or, finally, molecular and non-quantified (e.g. “Smith is tall and Jones is short”). We’ll now see that any sentence belonging to any given one of these categories can be (re)parsed to as to have the form ‹x has phi.› [409]

3.1 Non-quantified non-molecular sentences Any such sentence either says of some individual that it has some property or it says of some n-tuple says of two or more individuals or items that they are interrelated in a certain way (or, alternately, of some n-tuple that it has some n-place property). Thus, “Smith is tall” says of some individual that he has a certain property; and “Bob loves Sally” says that one individual (Bob) bears a certain relation (that of loving) with respect to some other individual (Sally), or, equivalently, that the ordered pair <Bob, Sally> has L, where L is a two-place property (a property of ordered pairs) such that for any x, y, ‹<x,y> has L› exactly if x loves y.

It’s obvious how to extend this procedure to sentences that affirm the existence of relations involving three or more objects. Let R* be a three-place predicate such that ‹<x,y,z> has R*”› is true just in case x is standing in between x and y. In that case, “Bob is standing in between Sally and Larry” is equivalent with ‹<Bob, Sally, Larry> has R*,” where ‹<x,y,z> has R*”› is true just in case x is standing in between x and y. ‹<Bob, Sally> has L› and ‹<Bob, Sally, Larry> has R*› both have the form ‹x has phi›, just like “Sally is tall.” Given these points, it’s clear, for any finite n, how to deal with sentences involving the existence of n-place relations.

3.1 Molecular, non-quantified sentences Let "K" be defined as follows: for any sentences S_1 and S_2, ‹<S_1, S_2> has K› is true iff the state of affairs described by S_1 is a consequence (of some kind or other) of the state of affairs described by S_2. Thus, "Smith broke his leg because he fell out of a tree" is equivalent with ‹<Smith broke his leg, Smith fell out of a tree.> has K›, which has the form ‹x has phi.› "Smith doesn't smoke" is equivalent with "the proposition that Smith smokes is false," which clearly has the form: ‹x has phi.› Other compound sentences are to be dealt with similarly.

3.2 Quantified generalizations A "quantified generalization" is any statement that says how many members one class of objects has in common with some other class of objects. "Some person smokes" is a quantified generalization; for it says that the class of people has at least one member in common with the class of smokers.

Bearing this in mind, let "E" be defined as follows: for any properties P and Q, ‹ <P, Q> has E› is true iff the class of things having P has a least one member with the class of things having Q. In that case, "some person smokes" is equivalent with ‹<the property of being a person, the property of being a smoker>, has E›, which obviously has the form ‹x has phi›. Other quantified generalizations are to be dealt with similarly.[410]

Since any given sentence is either atomic, quantified (and molecular), or non-quantified and molecular, we have thus established that any given sentence S is equivalent to some sentence having the form ‹x has phi.› Given this fact, there is an obvious answer to the question "what is it for a sentence to be meaningful?" A sentence is meaningful if it attributes some property to some object. A sentence S is meaningful if, for some object x and some property phi, S says that x has phi. In other words, a sentence is meaningful if it says anything about anything.

3.3 A note on "the absolute is one" and other supposed examples of meaningless *bona fide* sentences

The sentence-like expression

(i) The absolute is one

is presumably supposed to attribute some property to some parably important object (which, for expository reasons, we'll assume to be the universe.) Since this property isn't identified, (i) contains an undefined term and therefore says nothing. Once that term is defined, a meaningful statement results.

(i) is comparable to "x is tall." The reason "x is tall" says nothing is that "x" is undefined—it hasn't been assigned a referent. And no sooner is a referent assigned to "x" than "x is tall" becomes meaningful.

Before "x" is assigned a referent, "x is tall" is neither confirmable nor tautologous. But that's only a symptom of the real problem: *"x" doesn't have a referent*. The same thing *mutatis mutandis* is true of (i) and, indeed, of each of the logical positivists' other paradigms ("the nothing nothings," "the all is one," etc.).[411]

4.0 Russell's Paradox: McCready vindicated? Let K be the class of all classes that are not members of themselves---that don't have the property of being self-members.

(K_1) "K is a self-member"

entails, and is entailed by,

(K_2) “K is not a self-member.”

By the Law of Excluded Middle (LEM), *at least* one of those statements is true. By the Law of Non-Contradiction, *at most* one of those statements is true. But if one is true, so is the other, in which case LCN is violated.

Taking it for granted that K_1 is a *bona fide* sentence, and holding for the reasons just given that it encodes no proposition, Russell holds that K_1 is a true case of a sentence that is meaningless. Of course, if Russell is right, so is McCready.

But, although some of what Russell says is correct, the parts that are relevant to McCready’s thesis are not.

Here we must add to some points made earlier about the concept of (sentence-level) meaningfulness. First we must distinguish sentence-types from sentence-tokens.[412] The sentence-type:

(TE[413]) “That elephant doesn’t like that person”

is neither true nor false. But it’s meaningful. For, given any context C and given any token t of t that occurs in C, TE assigns truth-conditions to t. Those conditions are assigned by a rule similar to, if not coincident with, the following:

(SE[414]) If, in a given context C, there is exactly one salient elephant X and exactly one salient person Y, then an utterance in C of TE is true exactly if the proposition *X doesn’t like Y* is true. And if, in C, there isn’t a uniquely salient elephant or a uniquely salient person, then a token of TE in C fails to

encode a proposition or, therefore, to be true or false.

The meaning of a sentence-type is a rule that assigns truth-conditions to tokens of that type. [415]

Supposing this correct, the meaning of a sentence-*type* isn't a proposition; nor is it a set of truth-conditions. It is a function or rule that assigns truth-conditions to tokens of itself. [416]

t, we will suppose, is a token of TE that occurs in a context C where there is no salient elephant or person. You hear t. You don't know whether t is true or not. (You don't have good vision and thus can't tell whether there is a uniquely salient elephant or a uniquely salient person or *a fortiori* whether there is any animosity on the part of some such elephant towards some such person.) You know how to go about finding out whether t is true. You know that there must be one salient elephant, one salient man, etc. So it's because you *understand* t---it's because, in other words, you know what it means---that you know how to figure out whether it's true or not. (If you didn't speak English (and you didn't have some ulterior source of t-relevant information, e.g. a translator), you wouldn't understand t--- you wouldn't know what it meant, in other words---and you therefore couldn't set out to determine whether it was true or not.

It's worth noting that imperatives, interrogatives, and non-indicatives generally, don't have *truth*-conditions, even though they're meaningful, and instead have conditions of satisfaction. So being true (or false), though sufficient for having sentence-token-level meaning, is not necessary for it. But, what is more important in this context, it isn't necessary for it even if we focus only on indicative sentence-tokens. A sentence-token must have truth-conditions *in order* to affirm a proposition.

This argument doesn't involve a failure to distinguish type-meaning from token-meaning. t's meaning does *not* coincide with TE's meaning. t has truth-conditions. TE does not. TE's meaning is a rule (that assigns truth-conditions to its tokens). TE's meaning has nothing to do with this or that specific context or this or that specific utterance. A sentence that accurately says what TE means won't mention C or any other *specific* context. A sentence that accurately says what t means *will* mention C. So, in saying that t *is* meaningful, despite failing to encode a proposition, I'm not making the mistake of assigning it the meaning of the (very clearly meaningful) corresponding sentence-type.

This is not to deny the truism that, on at least one disambiguation of the term "meaning," the meaning of a token of "snow is white" is the proposition that snow is white. It is say that there are two kinds of sentence-token-meaning. Supposing that, in C, there is a uniquely salient elephant X and a uniquely salient person Y, one element of t's (two-tiered) semantic structure is given by identifying the proposition meant by t. The other is given by identifying t's truth-conditions. It is important to bear in mind that those truth-conditions are not identical with that proposition, since, holding the rules of English constant, the conditions that t must satisfy to be true are invariant with respect to whether t encodes a proposition.

4.1 Mapping this onto McCready's contention Let k_1 be a token of K_1. For the reasons given earlier, k_1 cannot possibly express a proposition. But k_1 has truth-*conditions.* To be sure, it's logically impossible that those conditions should be satisfied. But it's one thing to say that an expression has *no* truth-conditions, and it's quite another to say that it *does* have such conditions but that they cannot be fulfilled. And, very clearly, the semantic rules of English being what they are, k_1 does have truth-conditions, these

being given by the sentence:

(*) If there exists exactly one class K such that any given class C belongs to K iff C doesn't belong to itself, then k_1 is true iff K is a member of itself. [417]

Many conditions cannot possibly be fulfilled. The condition that a thing must fulfill to be an object that isn't identical with itself cannot be fulfilled. But that condition exists. That's why we can say that the class of things fulfilling it is empty.[418]

It would be the height of methodological insanity to say that, should the truth-conditions of the tokens of a certain sentence-type prove to be unsatisfiable, those tokens must be judged meaningless. Not all of the answers to the puzzles of philosophy, logic, and mathematics are, or ought to be, coded into the semantic rules of English (or any other language). We don't want our means of communication prejudging whether there are infinitely many primes, or perpetual motion machines, or square circles. We'd never acquire a means of communicating if we held off on adopting such a means *until* we'd established that none of the expressions belonging to it had necessarily unsatisfiable truth-conditions or affirmed necessarily false propositions. (And, in point of fact, the semantic rules of the English language *don't* prejudge such matters: those rules are not a repository for logical omniscience; they arose organically, in consequence of the thoughts and deeds of people most of whom were not great logicians.) Also, as previously stated, a language purged of all such tokens would *ipso facto* be drained of a good deal of its expressive power. It's only *because* there exist tokens like k_1 that we can express the truth that there is no class whose membership includes all and only non-self-inclusive classes. We need the

absurd to express, and often to establish, the non-absurd. (E.g. if we to establish (as Cantor did) that there are more reals than rationals, we need the absurd proposition that the reals can be put into a one-one correspondence with the rationals.)

5.0 Generalizing the points made in 4.0-4.1 Let tf be a token of the sentence-type:

(TF) “This sentence is false.”

According to the conventional wisdom:

(CW). tf is false if it’s true and true if it’s false. So, if it’s either true or false, it’s *both* true *and* false. Which is impossible. So it’s neither. Therefore there is no proposition that it express. Therefore, it is meaningless. But---it is usually said---it is a *bona fide* sentence-token.[419]

There two big problems with tf. Right now we’ll focus on the one that is most directly relevant to McCready’s thesis.

5.1 The first problem with tf Given only that there is no proposition affirmed by tf, it doesn’t follow that it’s meaningless. Indeed, tf is quite definitely meaningful. For it has truth-conditions. There is a certain object---a certain deposit of ink, in fact, just a few inches from your nose--such that tf is true just in case that ink-deposit is false. (That ink-deposit is, of course, none other than tf itself.) If that condition is fulfilled, tf is true. As a matter of logic, that condition cannot be fulfilled. But that’s irrelevant. The English language makes it crystal-clear how the world *would* have to be *if* tf were to

be true; it says what condition the world *would* have to fulfill.

If tf were actually meaningless, CW wouldn't go through. CW is a *reductio ad absurdam.* It has the form: "*If* tf affirms some proposition *p*, *then*, for some proposition *q*, *q and not-q* follows." In other words, it has the form: given what tf *means*, there is no way that it can be either true or false without being both; so there is no way it can be either.

Any sentence-token that can serve as the starting point of a *reductio ad absurdam*---indeed, any sentence-token has a meaning if it[420] can serve as any part of any cogent argument.

Most of CW is correct. tf does not express a proposition, since, just as CW says, any such proposition would violate LNC. Where CW goes wrong is in equating failure to affirm a proposition with lack of (sentence-token-level) meaningfulness.

To be sure, there is a difference between TF, on the one hand, and

(SC) "there are square circles,"

on the other. A token t* of SC *does* affirm a proposition, whereas tf does not. But both tokens are meaningful.

Tokens of SC aren't *semantically* false. Linguistic rules do no more in the way of settling whether tokens of SC are true than they do in the way of settling whether tokens of "dogs don't bite are true." Semantics only supplies the meanings; mathematics and zoology supply the truth-values.[421] But tokens of TF *are* semantically false. They are false entirely by virtue of what they mean---not by virtue of what they mean *plus* facts, however obvious, of logic or mathematics. Given *only* what tf means, it cannot be true without being false. The semantic rule that assigns it meaning is to the effect that it's

false if true. So the difference between tf and t* isn't the difference between meaningless and meaningful; it's the difference between a strictly semantics-based based failure to be true and a (partly) non-semantics-based failure to be true.

5.2 The second problem with tf (along with a correlative conjecture concerning self-reference-related paradoxes)

It is in virtue of what a given sentence-token *affirms* that it is true or false. Describing a sentence-token as "true" ("false") is an abbreviated way of saying that it affirms a true (false) proposition.[422] "This *sentence* is true (false)" really means "this *proposition* is true (false)." Suppose that the occurrence of "this sentence" in a token of "this sentence is true" refers to "snow is white." In that case, what is being said is an abbreviation for:

(TP) "The *proposition* meant by this sentence is false."

The semantic rule for TP is:

(STF) If, in context C, there is some uniquely salient proposition P, then a token t of TF is true exactly if P is true, and t is neither true nor false, as it fails to affirm anything, if there is no such proposition in C.

The proposition (supposing there to be one) affirmed by a given sentence-token is a function of (among other things) what its parts refer to.

This is a consequence of the "principle of compositionality"[423] namely:

(CM) What is meant (referred to) by a given expression is determined by what its parts meant (refer to)."[424]

Let O be the occurrence of "the proposition meant by this sentence" in a token t of TP. If O refers to nothing, t is neither true nor false. What t affirms depends on what O refers to. O must have a referent *independently* of what t affirms. Otherwise t won't affirm anything. This is easily shown. I'm having an exchange with my friend Jerry. I say: "*He* is *her* father." Since the context doesn't supply either one of the italicized terms with a referent, Jerry asks me who "he" refers to. I say "*her* father." Jerry then asks me who "*her*" refers to. I say "*his* daughter." If the answers I've given to Jerry are the only viable ones, then there is *no* viable answer to either question, in which case neither italicized terms refer to anything and I've therefore affirmed nothing. In an email exchange with Jerry, I say: "*the proposition meant by this sentence* is false." Jerry asks: "Which proposition is that?" I say: "The one meant by the underlined sentence." Jerry then asks: "Which proposition is meant by the underlined sentence?" I say: "the one meant by the italicized expression." Echoing what we just said, if the answers I've given are the only viable ones, then there is *no* viable answer to either question, in which case neither the underlined term nor the italicized terms refer to anything and I've therefore affirmed nothing.

The moral of this story: *Tokens of TP are neither true nor false.* So such tokens are not both true *and* false. So, assuming the correctness of the Strawson-Kaplan view of indexicals, the semantic rules of English *do not*

permit the construction of sentences that violate LNC.

Similar points hold with respect to:

(K_1) “K is a self-member.”

K is defined as “the class of all classes that don’t contain themselves.” Given this fact, there are, setting aside irrelevant nuances, two possibilities as to the identity of the semantic rule for “K”: rule is for semantic rule for “K” must be one of the following:

First possibility (P_1): If there exists some object O such that O is uniquely a class of all classes that don’t contain themselves, then a token of ‹…K…› affirms a proposition to the effect that:…*O*…; and if there is no such object, then such a token fails to affirm any proposition.

Second possibility (P_2): ‹…K…› affirms a proposition to the effect that: there exists exactly one class C that contains every class that doesn’t contain itself; moreover,…C…

P_1 and P_2 correspond to, respectively, Strawson’s and Russell’s analyses of definite descriptions.[425] If P_1 is the correct option, then (a given token of) “K is a self-member” is neither true nor false, and it therefore isn’t both true *and* false. If P_2 is the correct option, then (a given token of) “K is a self-member” isn’t both true *and* false---it’s just false. For, as

Russell himself shows, there cannot possibly be a class of the relevant sort, since, if there were, LNC would be violated.[426] In neither case do we have a sentence that is both true *and* false. So the Theory of Types (TT) is unnecessary, at least to the extent that the justification for it lies in K_1's violating LNC.

Russell (1903) believed K_1 to be a counterexample to the Axiom of Comprehension (namely, for any property P, there is a class C such that C contains every instance of P and nothing else*)*. This contention of his is generally accepted. But it's false. Some background points are needed to make it clear why this is so. The statement:

(S_1) "The class of all spoons has the property of not being member of itself"

is merely an abbreviation for:

(S_2) "The class of spoons isn't a spoon."

And the statement:

(T_1) "The class of non-spatiotemporal entities has the property of being member of itself"

is short for:

(T_2) "The class of non-spatiotemporal entities is non-spatiotemporal."

In general, when used meaningfully, the expression "property of (not) being a self-member" is eliminable. Why? Because "itself" isn't to be defined denotatively; it isn't to be defined by identifying its referent. That expression *per se* no more has a referent than any other pronoun. "He," the expression-type, has no referent; it is *occurrences* of "he" that have referents (cf. "John has a lot money, but *he* doesn't have a car"). Similarly, "the property of (not) being identical with itself" does not, in and of itself, have a referent; its *occurrences* have referents. The occurrences of it in S_1 and T_1 refer, respectively, to the property of being a spoon and the property of being non-spatiotemporal. (S_2 is equivalent with "the class of spoons doesn't have the property of being a spoon," the same thing *mutatis mutandis* holding of T_2.)

With these points in mind, let's take another look at:

(K_1) "K is a self-member (i.e. K has the property of being a member of itself)"

The property a thing must have to be a member of the class of spoons is that of being a spoon. What is the property a thing must have to be a member of K? In other words, what does "itself" refer to in K_1? "The property of *not* being a member of itself," it will be said. "Remember: K is the class of all classes that aren't members of themselves. So to say that K belongs to itself is to say that K doesn't belong to itself."

But this answer is brazenly circular. We wanted to know what "the property of being a member of itself" referred to, and the answer we're given is: "the property of *not* being a member of itself," which leaves us with the same question (*mutatis mutandis*) as before.

To clarify: We ask: “what property a thing must have to belong to K?” We ask, in other words: “what does ‘the property of being a member of K’ refer to?” We are told: “the property of not being a member of itself.” We then ask: “what does ‘the property of not being a member of itself’ refer to?” We are told: “the property of being a member of K.” Our questions haven’t been answered; we’re in the same position as Jerry (cf. the fictitious dialogue above).

Thus, “the property of (not) being a member of itself” *doesn’t refer to anything.* It’s a free variable; it’s like the occurrence of “he” in a decontextualized utterance of “he has no friends” (by which I mean an utterance in which neither the context nor any accompanying verbiage assigns a referent to “he”). *Therefore K_1 doesn’t express a proposition.* Since it doesn’t express a proposition, it isn’t true or false. Therefore, it isn’t both true and false. So we have no counterexample to the Axiom of Comprehension in K_1. Thus, to the extent that the motivation for TT lies in the contention that K_1 *is* such a counterexample, there is no motivation for TT.

Notice that the rules of English semantics are not at fault here. Those rules are unambiguous about how anaphoric terms are to be interpreted. Those rules make it clear that “x is a mad man” affirms no proposition, since it contains a free variable, and they make it clear that, for essentially the same reason, sentence(-tokens) containing free occurrences of prounoun---e.g. decontextualized tokens of “he is a mad man”---*do not affirm propositions.* So the semantic rules constitutive of the English language assign neither true *nor* falsity to K_1; so they don’t assign *both* truth *and* falsity to K_1. So they *don’t* countenance any violation of LCN, at least not to the extent that, relative to them, K_1 violates some logical principle. So those semantic rules *do not*, to that extent at least, need to be

modified in accordance with TT or otherwise logically reconstructed. So TT is unnecessary.

TT is also false. TT says that no meaningful sentence concerns *every* sentence. But TT is given by a sentence that violates that very stricture. For it's given by the sentence: "All meaningful sentences concern no sentences other than those that belong to orders (or types) lower than those to which they themselves belong." So TT defeats itself.

At the same time, even though no proposition is affirmed by K (or, strictly speaking, by its tokens), K is meaningful. For it has truth-conditions. For reasons of logic, those conditions cannot be fulfilled. (They could be fulfilled only if viciously circular arguments could be cogent. The fact that such arguments cannot be cogent isn't a by-product of our, or any other species', semantic rules; it's a matter of semantics-transcendent (though, of course, semantics-relevant) logical fact.)

Conclusion Every allegation that a given *bona fide* sentence(-token) was meaningless embodied one of three failures: first, a failure to distinguish between having no meaning, on the one hand, and having an idiotic meaning, on the other; second, a failure to distinguish between the property of affirming a proposition, on the one hand, and the property of having truth-conditions, on the other, it being the latter property, not the former, that is to be identified with the property of being meaningful; a failure to distinguish between type-meaning and token-meaning. Further, it can be shown that a *bona fide* sentence cannot possibly be meaningless. Any given sentence, no matter what its syntactic structure, is equivalent to one of the form ‹x is phi›, for some x and some property phi. If either ‹x› or is ‹phi› a nonsense term, like "tove" or "slithy," then ‹x is phi› is but a pseudo-sentence, like "the slithy toves did gyre and gimble in the wabe." If

‹x› is an empty referring term (e.g. a token of “that elephant” that occurs in an elephant-free environment), then (in such an environment) ‹x is phi›, though not affirming a proposition, has truth-conditions and is therefore meaningful. (Let t be a token of “that elephant is a menace” that occurs in an elephant-free environment. The semantic (meaning-assigning) rules of English being what they are, t *would* be correct *if* there a (uniquely) salient elephant and, moreover, any such elephant were a menace. t *is* true in any world W semantically just like ours in which, in the context in question, there is a (uniquely) salient elephant x and, moreover, x is a menace. Those clearly correct claims would be false if t had no meaning. It makes no sense to say that, *given* what the semantic rules of English are---in other words, *given* what the rules are that assign meaning to t---t has no meaning.)[427]

Addendum: The fact that “itself” (and other terms of its ilk, e.g. “herself,” “self-identical”) are always eliminable is a consequence of a more general principle, namely: The word “self” is *always* eliminable. In fact, *all* pronouns are eliminable. This point needs to be clarified and qualified. None of the qualifications undermines our critique of Russell’s analysis, but each should be stated for the sake of completeness.

Let’s start with a purely clarificatory point. The word “self” has two meanings or functions. Sometimes it can function as a garden-variety noun. That’s how it’s functioning in “Hume denied that he had a self.” And sometimes “self” functions as a way of constructing ‘terms of laziness’---expressions that are useful, because conducive to expedience, but technically eliminable. That’s how it functions in

(J_1) John punched himself,

which is short for

(J_2) John punched John.

When functioning denotatively (as in "my self (in other words, my soul or psychological essence, or some such) is something that underlies and unifies the various fleeting occurrences and states constitutive of my consciousness") isn't eliminable (at least not in any relevant sense). But (so I am trying to show) when functioning *anaphorically*, it *is* eliminable (cf. J_1 and J_2). Henceforth, all references to the term "self" are to the anaphoric device.

Now the requisite qualifications of our contention (that "self" is always eliminable): First, when the word "self" occurs in contexts governed by terms denoting propositional attitudes, e.g. "believes," "hopes"), it is not eliminable, at least not in any straightforward way. For example,

(L_1) "Larry believes himself to be a great pianist,"

isn't shorthand for:

(L_2) "Larry believes John to be a great pianist."

Larry might have amnesia and not *know* that he's Larry but still believe that Larry---with whom, unbeknownst to himself, he is identical--- to be a great pianist. The reason L_1 and L_2 aren't equivalent is that, in L_1, "self" falls within the scope of an expression ("believes") that denotes a propositional attitude. But in K_1 "self" doesn't occur within the scope of "believes" or,

indeed, *any* intensional operator: it is functioning extensionally. So this qualification is irrelevant to our critique of Russell.

Second, when a pronoun functions as a bound-variable, as opposed to a *bona fide* device of reference, it is not eliminable, at least not in any straightforward way. The sentence:

($\$_1$) "If someone gives you $1,000,000 and asks for nothing in return, that person has a good heart."

is not equivalent with:

($\$_2$) "If someone gives you $1,000,000 and asks for nothing in return, someone has a good heart."

That's because ($\$_1$)'s meaning is given by a sentence in which "that person" is replaced by a bound variable:

($\$_3$) "for any person x and anything y, if x gives y $1,000,000 and asks for nothing in return from y, then x has a good heart,"

But this is irrelevant to our critique of Russell's point, since, in K_1, "itself" clearly doesn't occur as a bound variable. K_1 is not a quantified generalization; it is a singular statement. It says, of some (putative) object K, that K isn't a member of K.

Finally, there are differences, other than strictly acoustical (or orthographic) ones, between J_1 andJ_2; but those differences are strictly pragmatic; they have nothing to do with any *semantic* differences between those two sentences. In J_2, it's not clear whether the two occurrences of

"John" co-refer; and, were it not for the suffix "self" or some other comparable device, someone uttering J_2 would have to know John's last name (which wouldn't necessarily help, since "John Smith," no less than "John" might be ambiguous) or or use some entirely *ad hoc* measure (such as inflecting his voice suggestively or, if he's writing, using sub-scripts) to make it clear that both occurrences of "John" co-referred. But these differences between J_1 and J_2, it must be stressed, are strictly pragmatic. For *given* that both occurrences of "John" are not mere homonyms---*given*, in other words, that there is some *one* semantic rule that assigns meaning to both of them ---J_1 and J_2 have the same semantics. So, while someone hearing J_1 may, for the reasons just given, have trouble knowing that they have the same semantics, that fact, it follows tautologically, is not itself a matter of semantics. But this is irrelevant to our critique of Russell, since, in alleging that K counterexamples the Axiom of Comprehension, Russell obviously isn't make a point that depends on the pragmatics of K or any other sentence.

.

Ayer, Alfred Jules. 1935. Language, Truth, and Logic. London: Victor Gollancz, LTD.

Ayer, A.J. 1959. "Verification and Experience." In Ayer, A.J. ed. *Logical Positivism.* New York: Free Press, pp. 228-246).

Bach, Kent. 1987. Thought and Reference. Clarendon Press, Oxford.

Barwise, Jon and Perry, John. 1983. Semantic innocence and uncompromising situations. In: Martinich, A.P. (Ed.), The Philosophy of

Language. Oxford University Press, New York, pp. 392-404.

Brown, Bryson (2006). “On Paraconsistency.” In Jacquette, Dale. A Companion to Philosophical Logic. New York: Blackwell, pp. 628-650.

Bridgeman, P.W. (1953). “The Logic of Modern Physics.” In Feigle, H. and Broderick, M. eds. Readings in the Philosophy of Science. New York: Century-Crofts-Appleton pp. 123-128.

Carnap, Rudolph. (1953a). In Feigle, H. and Broderick, M. eds. “Formal and Factual Science.” Readings in the

Carnap, Rudolph. (1953b). “Testability and Meaning.” In Feigle, H. and Broderick, M. eds. Readings in the Philosophy of Science. New York: Century-Crofts-Appleton pp. 47-92.

Davis, Wayne. 2003. Meaning, Expression, and Thought. Cambridge: Cambridge University Press.

Dummett, Michael. 1973. Frege: Philosophy of Language. Harvard University Press, Cambridge, MA.

Frege, Gottlob. 1891. “On Function and Concept.” In: Beaney, Michael (Ed.), The Frege Reader. Malden, MA: Blackwell, pp. 130-148.

Frege, Gottlob. 1892. “On Sense and Nominatum.” In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 190-202.

Frege, Gottlob. 1915. “My Basic Logical Insights.” In: Beaney, Michael (Ed.), The Frege Reader. Malden, MA: Blackwell, pp. 322-324.

Hempel, Carl. G. 1959. “Psychology in the Physical Language.” In Ayer, A.J. ed. *Logical Positivism.* New York: Free Press, pp. 165-198.

Hempel, Carl. G. 1959. “The Empiricist Criterion of Meaning.” In Ayer, A.J. ed. *Logical Positivism.* New York: Free Press, pp. 82-107.

Hempel, Carl. G. (2008). “Empiricist Criteria of Cognitive Significance.” In Lange, Marc ed. New York: Blackwell, pp. 10-21.

Kaplan, David , 1975. Dthat. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 316-329.

Kaplan, David. 1979. On the Logic of Demonstratives. Journal of Philosophical Logic 8, 71-98.

Kaplan, David. 1989. Demonstratives. In: Almog, J., et al.(Eds.), Themes from Kaplan. Oxford University Press, New York, pp. 471-555.

Kaplan, David. 1989b. Afterthoughts. In: Almog, J., et al. (Eds.), Themes from Kaplan. Oxford University Press, New York, pp. 557-534.

Lange, Marc (ed.). (2007). The Philosophy of Science. Oxford: Blackwell.

Long, Peter. 2001. Logical Form and Grammar. New York: Routledge.

Neurath, Otto. 1959. “Protocol Sentences.” In Ayer, A.J. ed. *Logical Positivism.* New York: Free Press, pp. 199-208).

Priest, Graham. 2006. Doubt Truth to be a Liar. Oxford, New York. Carroll, Lewis. 2010. *Through the Looking Glass and What Alice Found There.* New York: Create Space Publications.

Priest, Graham (2006b). “Logicians Setting Together Contradictories: A Perspective on Relevance, Paraconsistency, and Dialetheism.” In Jacquette, Dale. A Companion to Philosophical Logic. New York: Blackwell, pp. 651-664.

Priest, Graham (2006). “Logicians Setting Together Contradictories: A Perspective on Relevance, Paraconsistency, and Dialetheism.” In Jacquette, Dale. A Companion to Philosophical Logic. New York: Blackwell, pp. 651-664.

Martin, Robert (ed.). (1984) Recent Essays on Truth and the Liar Paradox. Oxford: Clarendon.

Recanati, Francois. 1993. Direct Reference: From Language to Thought. Blackwell, Oxford.

Recanati, Francois. 2004. Literal Meaning. Cambridge University Press, Cambridge.

Reichenbach, Hans. (1953). “The Verifiability Theory of Meaning.” In

Feigle, H. and Broderick, M. eds. Readings in the Philosophy of Science. New York: Century-Crofts-Appleton pp. 93-102.

Russell's (1902) *Letter to Frege* appears, with commentary, in van Heijenoort 1967:124-125.

Russell, Bertrand. 1905. On Denoting. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 203-211.

Russell, Bertrand. 1908. *Mathematical Logic as based on the theory of types* in van Heijenoort (ed.). 1967:153-182

Russell, Bertrand. 1918. The philosophy of logical atomism. In: Marsh, R.C. (Ed.), Logic and Knowledge. Routledge, London, pp. 167-272.

Salmon, Nathan. 1986. Frege's Puzzle. The M.I.T. Press, Cambridge, MA.

Salmon, Nathan. 2005. Metaphysics, Mathematics, and Meaning: Philosophical Papers I. Oxford University Press, New York.

Salmon, Nathan. 2007. Content, Cognition, and Communication: Philosophical Papers II. Oxford University Press, New York.

Saul, Jennifer. 1997. Substitution and simple sentences. Analysis 57/2, 102-108.

Schlick, Moritz. 1959. "The Foundation of Knowledge." In Ayer, A.J. ed.

Logical Positivism. New York: Free Press, pp. 209-227).

Strawson, Peter. 1950. On Referring. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 219-234.

Whitehead, Alfred North and Russell, Bertrand. 1927. *Principia Mathematica to *56*. London: Cambridge University Press.

A Review of Erin McCready's Review of *Conceptual Atomism and the Computational Theory of Mind*

In *Conceptual Atomism and the Computational Theory of Mind* (2007, Amsterdam), I argued that a failure to take into account some basic facts about pragmatics had led to some very wrong semantic views; and I further argued that those wrong semantic views led to some very wrong views about the nature of mind, the nature of self-knowledge, and the nature of causal explanation. More specifically, I argued that those wrong semantic views had led to, or given credence to, t *Conceptual Atomism* (CA) and the *Computational Theory of Mind* (CTM).

According to CA, your concept of Abraham Lincoln---in other words, your ability to have thoughts about him---isn't in any divisible into, or constituted by, other concepts of yours; that concept is a representational atom. In general, any given concept (setting aside those that are "patently phrasal," e.g. *the big brown cow over on that hill*) is an atom: a self-contained, indivisible unit. According to CTM, mind is to brain what hardware is to software. What calculators do isn't a very mimicry of thought: it *is* thought.

Erin McCready published a review of my book in the present journal.

[428] While there can be no doubt that Dr. McCready is a fine scholar, his review seriously misrepresent the contents of my book. In some cases, McCready imputes views to me that are antithetical to those that it is my express objective to refute. (For example, McCready describes my book as an attempt to refute "semantic externalism," when, as the subtitle of the book makes clear (*A defense of Semantic externalism and of Content Internalism*), the very opposite is true! This was not a mere slip on McCready's part: his review is largely dedicated to establishing the defectiveness of my (non-existent) attempt to refute semantic externalism.) And McCready's arguments against my contentions, so far as he correctly identified them, involved transparent non-sequiturs and, contrary to what McCready himself claimed, didn't constitute outright demolitions of those contentions. Upon reading Dr. McCready's review, Jacob Mey, then the editor-in-chief of this journal, urged me to publish a thorough response (hereby submitted) to McCready's review, stating that, given the circumstances, it was appropriate that the normal length-requirements be waived.

This paper is divided into four parts. In Part I, I defend my (2007) assertion, which McCready denies, that any *bona fide* linguistic expression is *ipso facto meaningful*. In Part II, I outline my argument against CTM and show that, *pace* McCready, that argument makes no objectionable assumptions about the nature of syntax. In Part III, I outline my reasons, which McCready distorts, for holding that, given some defensible if not platitudinous truths about language-use, direct reference does a singularly good job of *accommodating* the very data that is traditionally thought to warrant its rejection.

In Part IV, I outline my (2007) theory of propositions, which, without saying why he does so, McCready simply dismisses. If correct, that theory substantiates many of the claims made in Parts I-III. At the same time, that

theory is independent of the claims made in those sections and is to be evaluated on its own terms.

Part I: Solutions to some semantic antinomies based on an analysis of the concept of meaningfulness

2.0 "Colorless green ideas sleep furiously" absurd, but meaningful Dr. McReady says that

(MS) "colorless green ideas sleep furiously"[429]

is a *bona fide* sentence that is meaningless. Apart from citing MS, McCready does little to substantiate his claim that there are meaningless *bona fide* sentences.

MS is absurd, not meaningless. MS can be taken to mean that there is nothing that is a colorless green dream that isn't *also* something that sleeps furiously. Thus read, MS is true (albeit vacuously so) and therefore meaningful). MS can also be taken to mean that there are colorless green dreams and, moreover, that any such thing sleeps furiously. Thus read, it's false for a number of reasons (e.g. any green thing *ipso facto* isn't colorless, dreams don't sleep, things that sleep don't do so furiously). Another possible reading is that a thing's being a colorless green dream disposes it for some reason or other (e.g. for reasons of mathematical or natural law) to be something that sleeps furiously. (As they are ordinarily understood, "triangles have three sides" and "metal expands when heated" are, respectively, to the effect that, for reasons of logic (or mathematics or whatnot) a thing can't be a triangle without having three sides and that, for reasons of natural law, a thing can't be metal if it doesn't expand when

heated." [430])

2.0: Absurdity vs. meaninglessness

"Aargdsahsdflkjh" is meaningless. It isn't absurd. In response to an utterance of it, one can't say: "that's absurd!" It cannot appear as the penultimate line of a legitimate indirect proof. By contrast,

(SC) "there are square circles"

is absurd, and one *can* describe an utterance of it as preposterous; it *can* occur as the penultimate line of an indirect proof; and, importantly, it can be embedded in true statements (e.g. "it's not the case that there are square circles," "if *per impossible* there were square circles, there would be objects that both had, and lacked, uniform curvatures"). SC is absurd *because* it's meaningful. It's absurd by virtue of the fact that there is a proposition P such that P is SC's meaning and such that P is inherently incapable of being true. [431]

Russell (1940)[432] claimed that:

(QP) "quadruplicity drinks procrastination"

is a *bona fide* sentence that is meaningless. (Quadruplicity is the property being a four-membered set; or so Russell says---rightly, we'll assume.) Russell's supposedly meaningless sentence is false, since the property of having four members, being non-spatiotemporal, cannot drink anything.

Being false, it's meaningful.[433]

Though probably driven in part by a failure distinguish between ridiculous meanings and non-meanings, Russell's belief that QP is meaningless was probably also motivated by his acceptance of the Theory of Types, which we'll discuss in Sections 4.0-5.2 and which, in agreement with most contemporary logicians, find to be false and indeed self-defeating (true if false and therefore false).

According to Ryle (1949),

(OU) ‹Oxford University is in place p,›

is meaningless, since Oxford University cannot, according to Ryle, have a location. (The reason for this, Ryle plausibly says, is that universities are individuated at least partly by their functional, and therefore not strictly by their geographical characteristics.)[434]

Ryle's reasoning defeats itself. Ryle says that OU doesn't have a location. So he says that, for any place p, OU doesn't p. So, by Ryle's own lights, OU is false, not meaningless.

According to Carnap (1959a),

(TS) "this stone is thinking of Vienna"

is literally meaningless. In making this claim, Carnap was making the now familiar mistake of confusing ludicrousness with meaninglessness. Since stones don't think, TS if false, not meaningless.

So far as CG, QP, OU, and TS can non-hyperbolically be described as "nonsense," it's because "nonsense" has ceased to mean "without sense," and

has come to mean “having an idiotic sense,” or some such.

2.1 The embedded clause test Each of these sentences can meaningfully occur as a proper part of larger, indisputably meaningful sentences, for example:

(MSR) “if colorless green ideas sleep furiously, it follows that there are things that are both colorless and green.”

MSR is a garden-variety counterlogical statement, like “if (*per impossibile*) there exist numbers p and q such that p/q have no common divisor and such that $(p/q)^2=2$, then p and q both have, and don’t have, a common divisor.” If the occurrence in MSR of CG were meaningless, MSR would be meaningless. Since MSR is meaningful, so is CG. Exactly similar points hold of QP, OU, and TS.

Some sentences are brazenly absurd. Others are absurd but not obviously so. We use obviously absurd ones, like CG, to root out the absurd ones that *seem* reasonable (“2 has a rational square root”), thereby preventing the latter from infiltrating and rotting out entire departments of thought.

2.2 A failure to distinguish between words and word-like non-words that responsible for a failure to distinguish sentences from sentence-like sentences In many lectures of his that I attended (as a UCLA undergrad), David Kaplan, a well known semanticist, insisted that the opening two lines of Lewis Carroll’s Jabberwocky[435] (“ Twas brillig, and the slithy toves/Did gyre and gimble in the wabe”) a tokened a *bona fide* sentence. Those two lines no more token a sentence than a man wearing a Richard Nixon mask is

Richard Nixon.) It is irrelevant that "brillig," "slithy," "tove," etc. are meant to bring to mind the words "brilliant," "slimy," "toad," etc.

In general, there are expression-like entities that, given their phonetic properties along with facts about the linguistic environments in which they occur, are easily mistaken for bona fide expressions. Given what the words flanking it are and what it sounds like, the occurrence of "blerko" in "sometimes a blerko can be a man's best friend" is likely to be mistaken for an actual word. But it has no meaning and is thus no word. As this last pseudo-sentence shows, a failure to distinguish *bona fide* subsentential expressions from pseudo-subsentential expressions leads to a failure to distinguish sentence-like non-sentences from *bona fide* sentences.

2.3 Words that, in certain contexts, aren't functioning *as* words and therefore, in said contexts, *aren't* words Oftentimes, a homonym of an actual word occurs in a context in which no semantic rule assigns any meaning to it and in which it therefore means nothing, even though the context in question is one that is otherwise just like a sentence. What occurs in "the absolute is one"[436] is not the word "absolute," but a mere homonym of that word. For the occurrence of "absolute" in "the absolute is one" to be a word, there would have to be some semantic rule that assigned it a meaning. Although there are semantic rules that assign meaning(s) to homonyms of that occurrence, that occurrence per se is assigned no meaning. Anything that is assigned no meaning is *ipso facto* not a linguistic expression. So "the absolute is one" is not a sentence, at least not until definite values are assigned to "absolute" and "one." Until then, the only difference between it and ‹the phi is psi›[437] is that, whereas "phi" and "psi" are used *exclusively* as variables, "absolute" and "one" are not used exclusively as variables. (A

related point is that, when “absolute” and “one” are functioning as variables, their ranges would seem to be circumscribed by the meanings they have when functioning as constants.)

2.4 An important meta-semantic corollary Tautologously, anything to which no semantic rule assigns a meaning *ipso facto* isn’t a linguistic expression. I am contradicting myself if tell you that “blerb” is a word of English and then proceed to tell you that there is no semantic rule of English that assigns meaning to it (or its tokens). Obviously “blerb” can be flanked by the same sort of verbiage as actual expressions-----cf. “Jerry is a real blerb”---and, as this very example shows, might, depending on what the verbiage is, be naturally taken to have this or that meaning (in this case, one to the effect that Jerry has some sort of negative attribute, I should think). But until a semantic rule assigns it meaning, it’s not an expression; and once it becomes meaningful---once, say, “blerb” becomes synonymous with “creep” (which could happen: cf. “nerd,” “dweeb”)---a semantic rule (viz. *for any x, ‹x is a blerb› is true iff x is a creep*) *ipso facto* assigns it a meaning.

The semantic rules of English do not assign a meaning or referent to the token of “absolute” in “the absolute is one.” So in that context, “absolute” isn’t functioning as a word and is, it would seem, functioning as, if anything, a (restricted) variable (a variable whose substituends are presumably restricted to expressions denoting things that are august and important in some way).

Israel Sheffler[438] tries to explain why:

(GW[439][440]) “Some vessels contain glubbified water” [441]

isn't "scientifically" or "empirically" meaningful, whereas

(HW) "Some vessels contain boiling water" [442]

is clearly meaningful.

The only solutions to this problem that Scheffler acknowledges are ones which take it for granted, first, that (GW) is a *bona fide* sentence, albeit a meaningless one and, second, that it's the occurrence of "glubbified" in GW renders that sentence unfit for verification, falsification, or confirmation and, says Sheffler, is for that reason a sentence-like non-sentence.

The actual answer to Sheffler's question ("why isn't GW a *bona fide* sentence?") is a lot more obvious than Scheffler's discussion suggests. "Glubbified" means nothing. There is no semantic rule of English that assigns a meaning to it. It's not that it's assigned an unscientific or unempirical meaning. It's simply assigned no meaning at all, that being why the would-be sentence in which it's embedded I no sentence at all.

Suppose there were some rule R of the English language to the effect that for x to be "glubbified" is for x to contain little demons, whose existence couldn't be verified or falsified or confirmed, that, when present in a body of water, rendered that water unfit for human consumption. In that case, GW would be meaningful. Useful to science? Probably not. Meaningful? Yes. Why? Because, if it weren't meaningful, we couldn't judge that, *given its meaning*, it was empirically unverifiable/unfalsifiable/confirmable.[443]

3.0 Which sentences are meaningful?[444] All sentences are *ipso facto* meaningful.[445] Here's why. S is a meaningful sentence if S says that some

object x has some property phi and is thus equivalent to a sentence of the form: ‹x has phi.› [446]

Any given sentence is either (i) non-molecular (e.g. “Smith is tall,” “Smith is afraid of Jones,” or (ii) a quantified generalization (e.g. “some people smoke”) or, finally, molecular and non-quantified (e.g. “Smith is tall and Jones is short”). We’ll now see that any sentence belonging to any given one of these categories can be (re)parsed so as to have the form ‹x has phi.› [447]

3.1 Non-quantified non-molecular sentences Any such sentence either says of some individual that it has some property or it says of some n-tuple says of two or more individuals or items that they are interrelated in a certain way (or, alternately, of some n-tuple that it has some n-place property). Thus, “Smith is tall” says of some individual that he has a certain property; and “Bob loves Sally” says that one individual (Bob) bears a certain relation (that of loving) with respect to some other individual (Sally), or, equivalently, that the ordered pair <Bob, Sally> has L, where L is a two-place property (a property of ordered pairs) such that for any x, y, ‹<x,y> has L› exactly if x loves y.

It’s obvious how to extend this procedure to sentences that affirm the existence of relations involving three or more objects. Let R* be a three-place predicate such that ‹<x,y,z> has R*”› is true just in case x is standing in between x and y. In that case, “Bob is standing in between Sally and Larry” is equivalent with ‹<Bob, Sally, Larry> has R*,” where ‹<x,y,z> has R*”› is true just in case x is standing in between x and y. ‹<Bob, Sally> has L› and ‹<Bob, Sally, Larry> has R*› both have the form ‹x has phi›, just like “Sally is tall.” Given these points, it’s clear, for any finite n, how to deal with sentences involving the existence of n-place relations.

3.1 Molecular, non-quantified sentences Let "K" be defined as follows: for any sentences S_1 and S_2, ‹<S_1, S_2> has K› is true iff the state of affairs described by S_1 is a consequence (of some kind or other) of the state of affairs described by S_2. Thus, "Smith broke his leg because he fell out of a tree" is equivalent with ‹<Smith broke his leg, Smith fell out of a tree.> has K›, which has the form ‹x has phi.› "Smith doesn't smoke" is equivalent with "the proposition that Smith smokes is false," which clearly has the form: ‹x has phi.› Other compound sentences are to be dealt with similarly.

3.2 Quantified generalizations A "quantified generalization" is any statement that says how many members one class of objects has in common with some other class of objects. "Some person smokes" is a quantified generalization; for it says that the class of people has at least one member in common with the class of smokers.

Bearing this in mind, let "E" be defined as follows: for any properties P and Q, ‹ <P, Q> has E› is true iff the class of things having P has a least one member with the class of things having Q. In that case, "some person smokes" is equivalent with ‹<the property of being a person, the property of being a smoker>, has E›, which obviously has the form ‹x has phi›. Other quantified generalizations are to be dealt with similarly.[448]

Since any given sentence is either atomic, quantified (and molecular), or non-quantified and molecular, we have thus established that any given sentence S is equivalent to some sentence having the form ‹x has phi.› Given this fact, there is an obvious answer to the question "what is it for a sentence to be meaningful?" A sentence is meaningful if it attributes some property to some object. A sentence S is meaningful if, for some object x and some property phi, S says that x has phi. In other words, a sentence is meaningful

if it says anything about anything.

3.3 A note on "the absolute is one" and other supposed examples of meaningless *bona fide* sentences

The sentence-like expression

(i) The absolute is one[449]

is presumably supposed to attribute some property to some parably important object (which, for expository reasons, we'll assume to be the universe.) Since this property isn't identified, (i) contains an undefined term and therefore says nothing. Once that term is defined, a meaningful statement results.

(i) is comparable to "x is tall." The reason "x is tall" says nothing is that "x" is undefined—it hasn't been assigned a referent. And no sooner is a referent assigned to "x" than "x is tall" becomes meaningful.

Before "x" is assigned a referent, "x is tall" is neither confirmable nor tautologous. But that's only a symptom of the real problem: *"x" doesn't have a referent*. The same thing *mutatis mutandis* is true of (i) and, indeed, of each of the logical positivists' other paradigms ("the nothing nothings," "the all is one," etc.).[450]

4.0 Russell's Paradox: McCready vindicated? Let K be the class of all classes that are not members of themselves, i.e. that don't have the property of being self-members. In that case,

(K_1) "K is a self-member"

entails, and is entailed by,

(K_2) “K is not a self-member.”

By the Law of Excluded Middle (LEM), *at least* one of those statements affirms a truth (a true proposition). By the Law of Non-Contradiction, *at most* one of those affirms a truth. But if one is true, so is the other, in which case a given proposition is both true and false, which is impossible.

Taking it for granted that K_1 is a *bona fide* sentence, and holding for the reasons just given that it encodes no proposition, Russell holds that K_1 is a true case of a sentence that is meaningless. Of course, if Russell is right, so is McCready.

But, although some of what Russell says is correct, the parts that are relevant to McCready’s thesis are not.

Here we must add to some points made earlier about the concept of (sentence-level) meaningfulness. First we must distinguish sentence-types from sentence-tokens.[451] The sentence-type:

(TE[452]) “That elephant doesn’t like that person”

is neither true nor false. But it’s meaningful. For, given any context C and given any token t of t that occurs in C, TE assigns truth-conditions to t. Those conditions are assigned by a rule similar to, if not coincident with, the following:

(SE[453]) If, in a given context C, there is exactly one salient elephant X and

exactly one salient person Y, then an utterance in C of TE is true exactly if the proposition *X doesn't like Y* is true. And if, in C, there isn't a uniquely salient elephant or a uniquely salient person, then a token of TE in C fails to encode a proposition or, therefore, to be true or false.

The meaning of a sentence-type is a rule that assigns truth-conditions to tokens of that type.[454]

Supposing this correct, the meaning of a sentence-*type* isn't a proposition; nor is it a set of truth-conditions. It is a function or rule that assigns truth-conditions to tokens of itself.[455]

t, we will suppose, is a token of TE that occurs in a context C where there is no salient elephant or person. You hear t. You don't know whether t is true or not. (You don't have good vision and thus can't tell whether there is a uniquely salient elephant or a uniquely salient person or *a fortiori* whether there is any animosity on the part of some such elephant towards some such person.) You know how to go about finding out whether t is true. You know that there must be one salient elephant, one salient man, etc. So it's because you *understand* t---it's because, in other words, you know what it means---that you know how to figure out whether it's true or not. (If you didn't speak English (and you didn't have some ulterior source of t-relevant information, e.g. a translator), you wouldn't understand t---you wouldn't know what it meant, in other words---and you therefore couldn't set out to determine whether it was true or not.

It's worth noting that imperatives, interrogatives, and non-indicatives generally, don't have *truth*-conditions, even though they're meaningful, and instead have *conditions of satisfaction*. (John Searle (1969) made this point and, so far as I know, coined the term "conditions of satisfaction.") So being

true (or false), though sufficient for having sentence-token-level meaning, is not necessary for it. But, what is more important in this context, it isn't necessary for it even if we focus only on indicative sentence-tokens. A sentence-token must have truth-conditions *in order* to affirm a proposition.

This argument doesn't involve a failure to distinguish type-meaning from token-meaning. t's meaning does *not* coincide with TE's meaning. t has truth-conditions. TE does not. TE's meaning is a rule (that assigns truth-conditions to its tokens). TE's meaning has nothing to do with this or that specific context or this or that specific utterance. A sentence that accurately says what TE means won't mention C or any other *specific* context. A sentence that accurately says what t means *will* mention C. So, in saying that t *is* meaningful, despite failing to encode a proposition, I'm not making the mistake of assigning it the meaning of the (very clearly meaningful) corresponding sentence-type.

This is not to deny the truism that, on at least one disambiguation of the term "meaning," the meaning of a token of "snow is white" is the proposition that snow is white. It is to say that there are two kinds of sentence-token-meaning. Supposing that, in C, there is a uniquely salient elephant X and a uniquely salient person Y, one element of t's (two-tiered) semantic structure is given by identifying the proposition meant by t. The other is given by identifying t's truth-conditions. It is important to bear in mind that those truth-conditions are not identical with that proposition, since, holding the rules of English constant, the conditions that t must satisfy to be true are invariant with respect to whether t encodes a proposition.

Strawson (1950) discovered that the term "sentence-meaning" is ambiguous: there is sentence-type-meaning and sentence-token-meaning. A sentence-type's meaning, according to Strawson, is a rule that assigns propositions to tokens of a given sentence-type, those propositions being the

meanings of those tokens. We have seen that the term “sentence-token-meaning” is ambiguous. The “meaning” of some sentence-token t may refer either to t’s truth-conditions or to the proposition that t affirms. These are different things, since t may have truth-conditions without affirming a proposition. For the reasons given earlier, a token’s truth-conditions do not coincide with the meaning of the corresponding sentence-type. Finally, given a sentence-token t that does affirm a proposition, t’s truth-conditions don’t coincide with the proposition it affirms. The truth-conditions of a token t of “that woman is taller than that man” are given by a sentence along the lines of:

(SWM[456]) If, in a given context C, there is exactly one salient woman x and one salient man y, then a token of “that woman is taller than that man” is to the effect that x is taller than y.

But, supposing that, in C, there is one contextually salient woman W and one contextually salient man M, the proposition borne by t is to the effect that:

(WM[457]) W is taller than M.

SM says that somebody is taller than somebody else; SWM doesn’t say this. SWM isn’t about people’s heights. It’s about the semantic rules involved in our describing people’s heights. SWM is about language; SM is not. Propositions≠sentence-token-meanings.

4.1 Mapping this onto McCready’s contention Let k_1 be a token of K_1. For the reasons given earlier, k_1 cannot possibly express a proposition. But

k_1 has truth-*conditions*. To be sure, it's logically impossible that those conditions should be satisfied. But it's one thing to say that an expression has *no* truth-conditions, and it's quite another to say that it *does* have such conditions but that they cannot be fulfilled. And, very clearly, the semantic rules of English being what they are, k_1 does have truth-conditions, these being given by the sentence:

(*) If there exists exactly one class K such that any given class C belongs to K iff C doesn't belong to itself, then k_1 is true iff K is a member of itself. [458]

It would be the height of methodological insanity to say that, should the truth-conditions of the tokens of a certain sentence-type prove to be unsatisfiable, those tokens must be judged meaningless. [459] Not all of the answers to the puzzles of philosophy, logic, and mathematics are, or ought to be, coded into the semantic rules of English (or any other language). We don't want our means of communication prejudging whether there are infinitely many primes, or perpetual motion machines, or square circles. We'd never acquire a means of communicating if we held off on adopting such a means *until* we'd established that none of the expressions belonging to it had necessarily unsatisfiable truth-conditions or affirmed necessarily false propositions.

And, in point of fact, the semantic rules of the English language *don't* prejudge such matters: those rules are not a repository for logical omniscience; they arose organically, in consequence of the thoughts and deeds of people most of whom were not great logicians.

Also, as previously stated, a language purged of all such tokens would

ipso facto be drained of a good deal of its expressive power. It's only *because* there exist tokens like k_1 that we can express the truth that there is no class whose membership includes all and only non-self-inclusive classes. In general, we need the absurd to express, the non-absurd.

5.0 Generalizing the points made in 4.0-4.1 Let tf be a token of the sentence-type:

(TF) "This sentence is false."

According to the conventional wisdom:

(CW). tf is false if it's true and true if it's false. So, if it's either true or false, it's *both* true *and* false. Which is impossible. So it's neither. Therefore there is no proposition that it express. Therefore, it is meaningless. But---it is usually said---it is a *bona fide* sentence-token.[460]

There two problems with tf. Right now we'll focus on the one that is most directly relevant to McCready's thesis.

5.1 The first problem with tf Given only that there is no proposition affirmed by tf, it doesn't follow that it's meaningless. Indeed, tf is quite definitely meaningful. For it has truth-conditions. There is a certain object---a certain deposit of ink, in fact, just a few inches from your nose--such that tf is true just in case that ink-deposit is false. (That ink-deposit is, of course, none other than tf itself.) If that condition is fulfilled, tf is true. As a matter of logic, that condition cannot be fulfilled. But that's irrelevant. The English language makes it crystal-clear how the world *would* have to be *if* tf were to

be true; it says what condition the world *would* have to fulfill.

If tf were actually meaningless, CW wouldn't go through. CW is a *reductio ad absurdam.* It has the form: "*If* tf affirms some proposition *p*, *then*, for some proposition *q*, *q and not-q* follows." In other words, it has the form: given what tf *means*, there is no way that it can be either true or false without being both; so there is no way it can be either.

Any sentence-token that can serve as the starting point of a *reductio ad absurdam*---indeed, *any* sentence-token---has a meaning if it[461] can serve as any part of any cogent argument.

Most of CW is correct. tf does not express a proposition, since, just as CW says, any such proposition would violate LNC. Where CW goes wrong is in equating failure to affirm a proposition with a lack of (sentence-token-level) meaningfulness.

To be sure, there is a difference between TF, on the one hand, and a token t* of

(SC) "there are square circles,"

on the other. A token t* of SC *does* affirm a proposition, whereas tf does not. But both tokens are meaningful.

Tokens of SC aren't *semantically* false. Linguistic rules do no more in the way of settling whether tokens of SC are true than they do in the way of settling whether tokens of "dogs don't bite" are true. Semantics only supplies the meanings; mathematics and zoology supply the truth-values.[462]

By contrast, tokens of TF *are* semantically false. They are false entirely by virtue of what they mean---not by virtue of what they mean *plus* facts, however obvious, of logic or mathematics. Given *only* what tf means, it

cannot be true without being false. The semantic rule that assigns it meaning is to the effect that it's false if true. So the difference between tf and t* isn't the difference between the meaningless and the meaningful; it's the difference between a strictly semantics-based based failure to be true and a (partly) non-semantics-based failure to be true.

5.2 The second problem with tf (along with a correlative conjecture concerning self-reference-related paradoxes)

It is in virtue of what a given sentence-token *affirms* that it is true or false. Describing a sentence-token as "true" ("false") is an abbreviated way of saying that it affirms a true (false) proposition.[463] If it affirms anything, a given token of "this *sentence* is true (false)" affirms that some *proposition* is false. Technicalities aside, "this sentence is true (false)" means "this *proposition* is true (false)."

Suppose that the occurrence of "this sentence" in a token of "this sentence is true" refers to the sentence "snow is white." In that case, what is being said is an abbreviation for:

(TP) "The *proposition* meant by this sentence is false."

The semantic rule for TP is:

(STF) If, in context C, there is some uniquely salient proposition P, then a token t of TF is true exactly if P is true, and t is neither true nor false, as it fails to affirm anything, if there is no such proposition in C.

The proposition (supposing there to be one) affirmed by a given sentence-token is a function of (among other things) what its parts refer to. This is a consequence of the "principle of compositionality"[464] namely:

(CM) What is meant (referred to) by a given expression is determined by what its parts meant (refer to)."[465]

Let O be the occurrence of "the proposition meant by this sentence" in a token t of TP. If O refers to nothing, t is neither true nor false. What t affirms depends on what O refers to. O must have a referent *independently* of what t affirms. Otherwise t won't affirm anything.

This is easily shown. I'm having an exchange with my friend Jerry. I say: "*He* is *her* father." Since the context doesn't supply either one of the italicized terms with a referent, Jerry asks me who "he" refers to. I say "*her* father." Jerry then asks me who "*her*" refers to. I say "*his* daughter." If the answers I've given to Jerry are the only viable ones, then there is *no* viable answer to either question, in which case neither italicized terms refer to anything and I've therefore affirmed nothing. In an email exchange with Jerry, I say: "*the proposition meant by this sentence* is false." Jerry asks: "Which proposition is that?" I say: "The one meant by the underlined sentence." Jerry then asks: "Which proposition is meant by the underlined sentence?" I say: "the one meant by the italicized expression." Echoing what we just said, if the answers I've given are the only viable ones, there is *no* viable answer to either question, in which case neither the underlined term nor the italicized terms refer to anything and I've therefore affirmed

nothing.

The moral of this story: *Tokens of TP are neither true nor false.* So such tokens are not both true *and* false. So, assuming the correctness of the Strawson-Kaplan view of indexicals, the semantic rules of English *do not* permit the construction of sentences that violate LNC.

Similar points hold with respect to:

(K_1) "K is a self-member,"

where K is defined as "the class of all classes that don't contain themselves."

Given this fact, there are, setting aside irrelevant nuances, two possibilities as to the identity of the semantic rule for "K": rule is for semantic rule for "K" must be one of the following:

First possibility (P_1): If there exists some object O such that O is uniquely a class of all classes that don't contain themselves, then a token of ‹…K…› affirms a proposition to the effect that:…*O*…; and if there is no such object, then such a token fails to affirm any proposition.

Second possibility (P_2): ‹…K…› affirms a proposition to the effect that: there exists exactly one class C that contains every class that doesn't contain itself; moreover,…C…

P_1 and P_2 correspond to, respectively, Strawson's (1950) and Russell's (1905) analyses of definite descriptions.[466] If P_1 is the correct option,

then (a given token of) "K is a self-member" is neither true nor false, and it therefore isn't both true *and* false. If P_2 is the correct option, then (a given token of) "K is a self-member" isn't both true *and* false---it's just false. For, as Russell himself shows, there cannot possibly be a class of the relevant sort, since, if there were, LNC would be violated.[467] In neither case do we have a sentence that is both true *and* false. So the Theory of Types (TT) is unnecessary, at least to the extent that the justification for it lies in K_1's violating LNC.

Russell (1902, 1903, 1908) believed K_1 to be a counterexample to the Axiom of Comprehension (namely, *for any property P, there is a class C such that C contains every instance of P and nothing else)*. This contention of his is generally accepted.[468] But it's false. Some background points are needed to make it clear why this is so. The statement:

(S_1) "The class of all spoons has the property of not being member of itself"

is merely an abbreviation for:

(S_2) "The class of spoons isn't a spoon."

And the statement:

(T_1) "The class of non-spatiotemporal entities has the property of being member of itself"

is short for:

(T_2) “The class of non-spatiotemporal entities is non-spatiotemporal.”

In general, when used meaningfully, the expression “property of (not) being a member of itself” is eliminable. Why? Because “itself” isn’t to be defined denotatively; it isn’t to be defined by identifying its referent. That expression *per se* no more has a referent than any other pronoun. “He,” the expression-type, has no referent; it is *occurrences* of “he” that have referents (cf. “John has a lot money, but *he* doesn’t have a car”). Similarly, “the property of (not) being identical with itself” does not, in and of itself, have a referent; its *occurrences* have referents. The occurrences of it in S_1 and T_1 refer, respectively, to the property of being a spoon and the property of being non-spatiotemporal. (S_2 is equivalent with “the class of spoons doesn’t have the property of being a spoon,” the same thing *mutatis mutandis* holding of T_2.)

With these points in mind, let’s take another look at:

(K_1) “K is a self-member (i.e. K has the property of being a member of itself)”

The property that a thing must have to be a member of the class of spoons is that of being a spoon. What is the property a thing must have to be a member of K? In other words, what does “itself” refer to in K_1?

“The property of *not* being a member of itself,” it will be said. “Remember: K is the class of all classes that aren’t members of themselves. So to say that K belongs to itself is to say that K doesn’t belong to itself.”

But this answer is brazenly circular. We wanted to know what “the

property of being a member of itself" referred to, and the answer we're given is: "the property of *not* being a member of itself," which leaves us with the same question (*mutatis mutandis*) as before.

We ask: "What property must a thing must have to belong to K? What does 'the property of being a member of K' refer to?" We are told: "the property of not being a member of itself." We then ask: "What does 'the property of not being a member of itself' refer to?" We are told: "The property of being a member of K." Our questions haven't been answered; we're in the same position as Jerry.

Thus, "the property of (not) being a member of itself" *doesn't refer to anything.* It's a free variable; it's like the occurrence of "he" in a token of "he has no friends" that occurs in a context in which nothing assigns any referent to the occurrence of "he." *Therefore K_1 doesn't express a proposition.* Since it doesn't express a proposition, it isn't true or false. Therefore, it isn't both true and false. So we have no counterexample to the Axiom of Comprehension in K_1. Thus, to the extent that the motivation for TT lies in the contention that K_1 *is* such a counterexample, there is no motivation for TT.

Notice that the rules of English semantics are not at fault here. Those rules are unambiguous as to how anaphoric terms are to be interpreted. Those rules make it clear that "x is a mad man" affirms no proposition, since it contains a free variable; and those rules make it clear that, for much the same reason, sentence(-tokens) containing free occurrences of pronouns *do not affirm propositions.* So the semantic rules constitutive of the English language assign neither truth *nor* falsity to K_1; so they don't assign *both* truth *and* falsity to K_1. So they *don't* countenance any violation of LCN, at least not to the extent that, relative to them, K_1 violates some logical principle. So those semantic rules *do not*, to that

extent at least, need to be modified in accordance with TT or otherwise logically reconstructed. So TT is unnecessary.

5.2.1 A positive argument against TT TT says that no meaningful sentence concerns *every* sentence. But TT is given by a sentence that violates that very stricture. For it's given by the sentence: "All meaningful sentences concern no sentences other than those that belong to orders (or types) lower than those to which they themselves belong." So TT defeats itself.

At the same time, even though no proposition is affirmed by K (or, strictly speaking, by its tokens), K is meaningful. For it has truth-conditions. For reasons of logic, those conditions cannot be fulfilled: for they could be fulfilled only if *per impossibile* viciously circular arguments could be cogent. But the fact that such arguments cannot be cogent isn't a by-product of our, or any other species', semantic rules; it's a matter of semantics-transcendent (though, of course, semantics-relevant) logical fact.

Conclusion to Part I We've considered several allegation to the effect that some expression was a both *bona fide* sentence(-token) and was also meaningless. Each such allegation, we found, embodied one of three failures: first, a failure to distinguish between having no meaning, on the one hand, and having an idiotic meaning, on the other; second, a failure to distinguish between the property of affirming a proposition, on the one hand, and the property of having truth-conditions, on the other, it being the latter property, not the former, that is to be identified with the property of being meaningful; a failure to distinguish between type-meaning and token-meaning.

Further, it can be shown that a *bona fide* sentence cannot possibly be

meaningless. Any given sentence, no matter what its syntactic structure, is equivalent to one of the form ‹x is phi›, for some x and some property phi. If either ‹x› or ‹phi› is a nonsense term, like “tove” or “slithy,” then ‹x is phi› is but a pseudo-sentence, like “the slithy toves did gyre and gimble in the wabe.” If ‹x› is an empty referring term (e.g. a token of “that elephant” that occurs in an elephant-free environment), then (in such an environment) ‹x is phi›, though not affirming a proposition, has truth-conditions and is therefore meaningful. (Let t be a token of “that elephant is a menace” that occurs in an elephant-free environment. The semantic (meaning-assigning) rules of English being what they are, t *would* be correct *if* there a (uniquely) salient elephant and, moreover, any such elephant were a menace. t *is* true in any world W semantically just like ours in which, in the context in question, there is a (uniquely) salient elephant x and, moreover, x is a menace. Those clearly correct claims would be false if t had no meaning. It makes no sense to say that, *given* what the semantic rules of English are---in other words, *given* what the rules are that assign meaning to t---t has no meaning.)[469]

Addendum: The fact that occurrences of “itself” is are always eliminable is a consequence of a more general principle, namely: the term “self” is *always* eliminable. (I am referring to the term “self” *qua* suffix, as in “himself,” not *qua* garden-variety noun, as in “people have selves, whereas lizards do not.” All references to “self” are to be thus disambiguated.[470]) In fact, *all* pronouns are eliminable. This point is subject to three qualifications, none of which redounds to the credit of Russell’s analysis of K_1.

First, when the word “self” occurs in contexts governed by terms

denoting propositional attitudes, e.g. “believes,” “hopes”), it is not eliminable, at least not in any straightforward way. For example,

(L_1) “Larry believes himself to be a great pianist,”

isn’t shorthand for:

(L_2) “Larry believes John to be a great pianist.”[471]

Larry might have amnesia and not *know* that he’s Larry but still believe that Larry---with whom, unbeknownst to himself, he is identical--- to be a great pianist. The reason L_1 and L_2 aren’t equivalent is that, in L_1, “self” falls within the scope of an expression (“believes”) that denotes a propositional attitude. But in K_1 “self” doesn’t occur within the scope of “believes” or, indeed, *any* intensional operator: it is functioning extensionally. So this qualification is irrelevant to our critique of Russell.

Second, when a pronoun functions as a bound-variable, as opposed to a *bona fide* device of reference, it is not eliminable, at least not in any straightforward way. The sentence:

($\$_1$) “If someone gives you $1,000,000 and asks for nothing in return, that person has a good heart.”

is not equivalent with:

($\$_2$) “If someone gives you $1,000,000 and asks for nothing in return, someone has a good heart.”

That's because ($\$_1$)'s meaning is given by a sentence in which "that person" is replaced by a bound variable:

($\$_3$) "for any person x and anything y, if x gives y $1,000,000 and asks for nothing in return from y, then x has a good heart,"

But this is irrelevant to our critique of Russell's point, since, in K_1, "itself" clearly doesn't occur as a bound variable. K_1 is not a quantified generalization; it is a singular statement. It says, of some (putative) object K, that K isn't a member of K.

Finally, there are differences, other than strictly acoustical (or orthographic) ones, between J_1 andJ_2; but those differences are strictly pragmatic; they have nothing to do with any *semantic* differences between those two sentences. In J_2, it's not clear whether the two occurrences of "John" co-refer; and, were it not for the suffix "self" or some other comparable device, someone uttering J_2 would have to know John's last name (which wouldn't necessarily help, since "John Smith," no less than "John" might be ambiguous) or or use some entirely *ad hoc* measure (such as inflecting his voice suggestively or, if he's writing, using sub-scripts) to make it clear that both occurrences of "John" co-referred. But these differences between J_1 and J_2, it must be stressed, are strictly pragmatic. For *given* that both occurrences of "John" are not mere homonyms. (Given that there is some *one* semantic rule that assigns meaning to both of them, J_1 and J_2 have the same semantics.) So, while someone hearing J_1 may, for the reasons just given, have trouble knowing that they have the same semantics, that fact, it follows tautologically, is not itself a matter of semantics. But this is irrelevant to our critique of Russell. For, in alleging

that K counterexamples the Axiom of Comprehension, Russell obviously isn't make a point that depends on the pragmatics of K or any other sentence.

Part II: The Computational Theory of Mind

xxx

.

Ayer, Alfred Jules. 1935. Language, Truth, and Logic. London: Victor Gollancz, LTD.

Ayer, A.J. 1959. "Verification and Experience." In Ayer, A.J. ed. *Logical Positivism.* New York: Free Press, pp. 228-246.

Bach, Kent. 1987. Thought and Reference. Clarendon Press, Oxford.

Barwise, Jon and Perry, John. 1983. Semantic innocence and uncompromising situations. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 392-404.

Brown, Bryson (2006). "On Paraconsistency." In Jacquette, Dale. A Companion to Philosophical Logic. New York: Blackwell, pp. 628-650.

Bridgeman, P.W. (1953). "The Logic of Modern Physics." In Feigle, H. and Broderick, M. eds. Readings in the Philosophy of Science. New York: Century-Crofts-Appleton pp. 123-128.

Carnap, Rudolph. (1953a). In Feigle, H. and Broderick, M. eds. “Formal and Factual Science.” Readings in the Philosophy of Science. New York: Century-Crofts-Appleton pp. 39-46.

Carnap, Rudolph. (1953b). “Testability and Meaning.” In Feigle, H. and Broderick, M. eds. Readings in the Philosophy of Science. New York: Century-Crofts-Appleton pp. 47-92.

Carnap, Rudolph. 1959a. “The Elimination of Metaphysics Through Logical Analysis of Language.” In Ayer, A.J. ed. *Logical Positivism.* New York: Free Press, pp. 60-81.

.

Carnap, Rudolph. 1959b. “Psychology in Physical Language.” In Ayer, A.J. ed. *Logical Positivism.* New York: Free Press, pp. 165-198.

.

Carroll, Lewis. 2010. *Through the Looking Glass and What Alice Found There.* New York: Create Space Publications.

Davis, Wayne. 2003. Meaning, Expression, and Thought. Cambridge: Cambridge University Press.

Dummett, Michael. 1973. Frege: Philosophy of Language. Harvard University Press, Cambridge, MA.

Frege, Gottlob. 1891. “On Function and Concept.” In: Beaney, Michael (Ed.), The Frege Reader. Malden, MA: Blackwell, pp. 130-148.

Frege, Gottlob. 1892. “On Sense and Nominatum.” In: Martinich, A.P. (Ed.),

The Philosophy of Language. Oxford University Press, New York, pp. 190-202.

Frege, Gottlob. 1915. “My Basic Logical Insights.” In: Beaney, Michael (Ed.), The Frege Reader. Malden, MA: Blackwell, pp. 322-324.

Gödel, Kurt. 1931. "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I", *Monatshefte für Mathematik und Physik*, 38: 173–198. Reprinted in Gödel 1986, pp. 144–195.

Gödel, Kurt. 1953. “Is Mathematics Syntax of Language.” Lecture manuscript. Printed in Gödel 1995. pp. 334–356.

Gödel, Kurt, 1986, *Collected Works. I: Publications 1929–1936*. S. Feferman, S. Kleene, G. Moore, R. Solovay, and J. van Heijenoort (eds.), Oxford: Oxford University Press.

Gödel, Kurt. 1995. Gödel, Kurt, 1995, *Collected Works. III: Unpublished essays and lectures*. S. Feferman, J. Dawson, S. Kleene, G. Moore, R. Solovay, and J. van Heijenoort (eds.), Oxford: Oxford University Press.

Grice, H.P. 1975. “Logic and conversation. In Cole, P and Morgan, J.L. (Eds.) Syntax and Semantics, Vol. 3: Speech Acts, pp. 41—58.

Hempel, Carl. G. 1959. “The Empiricist Criterion of Meaning.” In Ayer, A.J. ed. *Logical Positivism.* New York: Free Press, pp. 82-107.

Hempel, Carl. G. 2008. “Empiricist Criteria of Cognitive Significance.” In

Lange, Marc ed. New York: Blackwell, pp. 10-21.

Kaplan, David , 1975. Dthat. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 316-329.

Kaplan, David. 1979. On the Logic of Demonstratives. Journal of Philosophical Logic 8, 71-98.

Kaplan, David. 1989. Demonstratives. In: Almog, J., et al.(Eds.), Themes from Kaplan. Oxford University Press, New York, pp. 471-555.

Kaplan, David. 1989b. Afterthoughts. In: Almog, J., et al. (Eds.), Themes from Kaplan. Oxford University Press, New York, pp. 557-534.

Lange, Marc (ed.). (2007). The Philosophy of Science. Oxford: Blackwell.

Long, Peter. 2001. Logical Form and Grammar. New York: Routledge.

Martin, Robert (ed.). (1984) Recent Essays on Truth and the Liar Paradox. Oxford: Clarendon.

Neurath, Otto. 1959. “Protocol Sentences.” In Ayer, A.J. ed. *Logical Positivism*. New York: Free Press, pp. 199-208).

Perry, John. 1977. “The Problem of the Essential Indexical.” *Nous* 13(1), pp. 3-21.

Priest, Graham. 2006a. Doubt Truth to be a Liar. Oxford, New York. Carroll,

Lewis. 2010. *Through the Looking Glass and What Alice Found There.* New York: Create Space Publications.

Priest, Graham (2006b). “Logicians Setting Together Contradictories: A Perspective on Relevance, Paraconsistency, and Dialetheism.” In Jacquette, Dale. A Companion to Philosophical Logic. New York: Blackwell, pp. 651-664.

Quine, W.V.O. 1966. The *Ways of Paradox.* New York: Random House.

Recanati, Francois. 1993. Direct Reference: From Language to Thought. Blackwell, Oxford.

Recanati, Francois. 2004. Literal Meaning. Cambridge University Press, Cambridge.

Reichenbach, Hans. (1953). “The Verifiability Theory of Meaning.” In Feigle, H. and Broderick, M. eds. Readings in the Philosophy of Science. New York: Century-Crofts-Appleton pp. 93-102.

Russell, Bertrand. 1902. “Letter to Frege.” In van Heijenoort, J. (Ed.) 124-125.

Russell, Bertrand. 1903. Principles of Mathematics. New York: Norton.

Russell, Bertrand. 1905. On Denoting. In: Martinich, A.P. (Ed.), The Philosophy of Language. Oxford University Press, New York, pp. 203-211.

Russell, Bertrand. 1908. *Mathematical Logic as based on the theory of types* in van Heijenoort. J. (Ed.). 1967:153-182

Russell, Bertrand. 1918. The philosophy of logical atomism. In: Marsh, R.C. (Ed.), Logic and Knowledge. Routledge, London, pp. 167-272.

Salmon, Nathan. 1986. Frege's Puzzle. The M.I.T. Press, Cambridge, MA.

Salmon, Nathan. 2005. Metaphysics, Mathematics, and Meaning: Philosophical Papers I. Oxford University Press, New York.

Salmon, Nathan. 2007. Content, Cognition, and Communication: Philosophical Papers II. Oxford University Press, New York.

Saul, Jennifer. 1997. Substitution and simple sentences. Analysis 57/2, 102-108.

Schlick, Moritz. 1959. "The Foundation of Knowledge." In Ayer, A.J. ed. *Logical Positivism.* New York: Free Press, pp. 209-227).

Scheffler, Israel. 1981. The Anatomy of Inquiry. Cambridge, MA: Harvard University Press.

Searle, John. 1969. Expression and Meaning. Cambridge: Cambridge University Press.

Strawson, Peter. 1950. On Referring. In: Martinich, A.P. (Ed.), The

Philosophy of Language. Oxford University Press, New York, pp. 219-234.

Van Heijenoort, Jean. (Ed.). 1967. From Frege to Godel: A Source Book in Mathematical Logic, 1879-1931. Cambridge, MA: Harvard University Press.

The semantics and pragmatics of demonstrative expressions: What Kaplan failed to do

1.0 The concept of presemantics in relation to indexicals According to David Kaplan, demonstratives are "directly referential." (Examples of demonstratives are: "that guy" and "this apple." [472]) In other words, in virtue of having the form "…E…," where E is an occurrence of a demonstrative, a sentence-token S is true exactly if….O…

Kaplan's theory appears to have the counterintuitive consequence that sentence-tokens with radically different cognitive values can be identical in respect of their literal meanings. No real headway has been made on how to solve this problem.[473]

For the sake of argument, suppose that, as Kaplan (1989) holds, demonstrative expressions are "directly referential." In that case, there is some object x such that proposition literally meant by a token of:

(A) "that overbearing and generally unsavory man standing over there next to that atrocious painting is a professor of anthropology"

is simply:

(B) *x is a professor of anthropology.*

But obviously what is *communicated* by a token of (A) will not be so threadbare. What is communicated is (inter alia) that somebody is overbearing and generally unsavory and is also standing next to some painting that is atrocious. Thus, what is *communicated* by a token of (A) will involve concepts like *painting, overbearing, unsavory* that are quite absent from (B). How is this to be reconciled with Kaplan's thesis that (B) is the literal meaning of (A)?

There is no problem. We need only take into account the fact that people typically understand the expressions that they encounter *on the basis of their knowledge of the semantic rules that assign them meaning.* The semantic rule for the demonstrative expression "that overbearing and generally unsavory man standing over there next to that atrocious painting" is:

(C) For any context C, and any predicate phi, if somebody x in C is uniquely a salient overbearing and generally unsavory man standing next to a uniquely salient painting y, then a token in C of ⌜that overbearing and generally unsavory man standing over there next to that atrocious painting has phi⌝ means: *x has phi.*

The rule for (A) specifically is:

(D) For any context C, if there somebody x in C is a uniquely salient overbearing and generally unsavory man standing next to a uniquely salient

painting y, then a token in C of "that overbearing and generally unsavory man standing over there next to that atrocious painting is a professor of anthropology" means: *x is a professor of anthropology*.

When one hears a token of (A), one has to work *through* (D). Anyone who knows (D) will know that an utterance of (A) will not be true unless, in the context of utterance, somebody x is a uniquely salient overbearing and generally unsavory man standing next to a uniquely salient atrocious painting y, and *x* is a professor of anthropology. So even though what a token of (A) literally means is (B), and thus doesn't involve the concepts of being unsavory or over-bearing, what such a token *communicates* to someone does concern these very concepts. Because of the information that one must *work through* in order to assign the right proposition to a token of (A), what such a token communicates is very different from what it literally means.

I will refer to information conveyed in this way as "pre-semantic implicature." We will find that what we just said about tokens of (A) is true, to some degree or other, of *all* sentences. In connection with this, we will find that pre-semantic implicatures are exponentially more powerful than the post-semantic implicatures studied by Grice.

1.1 The solution to an apparent problem with our analysis There is a problem with the analysis just proposed. (We're assuming in this context that Kaplan's theory is correct.) To wit: An utterance of "that guy is evil" can have a radically different cognitively value from a different utterance of that same sentence, *even though both tokens of "that guy" refer to the same thing.* A story will both clarify this statement and suggest a way to solve the problem that it raises.

Somebody who is wearing a ski-mask, and who I therefore don't

recognize, deftly snatches my pocket from my wallet. As he's running off, I point at him and yell: 'that man is a thief!' Let U_1 be this utterance. There is some individual x such that x has just stolen my wallet and such that what I've just said is true exactly if x is a thief.

The next day, my lovable office-mate Steve eats one of the cupcakes that was on my desk. I jokingly point at him and say: 'that man is a thief.' Let U_2 be *this* utterance. There is some x such that x just ate my cupcake and such that U_2 is true exactly if: x is a thief.

Unbeknownst to me, Steve is the pick-pocket, and there is some individual x, namely Steve, such that each of U_1 and U_2 is true if and only if x is a thief. Thus, U_1 and U_2 have the very same literal meanings.[474] But I don't know this, even though I speak English perfectly and, on each occasion, obviously understand perfectly well what it is that I'm saying.

How this is possible? Our sense-perceptions describe things. My uttering U_1 was a response to my being given visual description of Steve. That description was to the effect that:

(i) there is some man x such that x is wearing a ski-mask and such that x is running off into the distance.

But U_1's literal meaning is not that that effect. There is some man x such that x is a wearing a ski-mask (etc.), such that in uttering U_1 I was saying that:

(ii) x is a thief.

The meaning of U_1, being identical with (ii), is quite threadbare. But I *grasped* that threadbare meaning through my descriptively rich visual

perception, whose content is given by (i).

My uttering U_2 was a response to my being given a different description of Steve. That description was to the effect that:

(iii) there is some man x such that x is a portly amicable fellow who is sitting over in that chair.

But U_2's literal meaning is not that that effect. There is some man x such that x is a portly amicable fellow (etc.) such that, in uttering U_2, I was saying that:

(ii) x is a thief.

Echoing what we said a moment ago, the meaning of U_2, being identical with (ii), is quite threadbare. But I *grasped* that threadbare meaning through my descriptively rich visual perception, whose content is given by (iii).

Because I grasped (ii) by way of different bodies of perceptual (descriptive) information, I didn't know, when uttering U_2, that what I was affirming the same thing I was affirming in uttering U_1. Oftentimes, literal meaning is cloaked by the *pre*semantic information through which it is grasped, and semanticists are needed to uncloak it.

So, given some sentence-token st containing an indexical, there are three things we must distinguish. First: What st literally means. (So if O is referent of "that thief" in a token of "that guy is a thief," the proposition affirmed by st is to the effect that O is a thief, and it is not to any other effect.)

Second, we must distinguish st's literal meaning from the semantic rules on the basis of a knowledge of which one understands st. In this context, the semantic rule for "that guy" is of special importance. Anyone who

understands expression-*type* “that guy”---anyone who knows (in the relevant, operational way in which speakers of a language can be said to know it) what the *character* is of “that guy”---*ipso facto* that knows a token of “that guy” won’t refer to an object O unless O is a male in the context in which that token occurred and who, perhaps because of an act of pointing on the speaker’s part, is salient. And by virtue of having that semantic knowledge and, more importantly, by virtue of having to exploit it to assign the right literal meaning to so---by virtue of these two things, st, though not *itself* containing the concepts *man* or *contextually salient object*----will inevitably (non-semantically) convey a message of which concepts are constituents. The concepts *man* and *contextually salient object* (and/or whatever other concepts are internal to the character of “that man”) are nowhere to be found st’s literal meaning. Where they *are* to be found is in the rules on the basis of one’s knowledge of which is a *pre*condition[475] for knowing what st (literally) means. So such concepts, such information, is appropriately characterized as presemantic.

Third, we must distinguish the sensory information through which st is presented to one in the first place. st didn’t just occur in a vacuum. It occurred in a certain context, having distinctive spatiotemporal, physical, social, and possibly cultural components. One’s sensory awareness of st is necessarily embedded in, or accompanied by, sensory awarenesses the information embodied in which has nothing to do with st’s semantics *or* with the sort of presemantic information discussed in the last paragraph. But, as we’ve seen (and are about to see again, this time in connection with proper names, as opposed to indexicals), this sensory information can dramatically effect the cognitive value of a given sentence. Since such sensory information is a prerequisite for even knowing of st’s existence, or *a fortiori* of knowing which semantic rules are the ones on the basis of which st’s literal meaning is

to be determined, this sensory information might be described as *pre*presemantic. But it's better, I think, to use the term "presemantic" to cover the totality of information---some of it consisting linguistic rules, some of it consisting of non-linguistic, purely sensory information. The important point is that both kinds of information, though each is distinct from st's literal meaning, can have an enormously important role in determining what st *communicates.*

Given a few platitudes about presemantics, most of which we've already stated, it's easy to solve some disputes concerning proper names that, despite decades of acrid disputes, have been quite fruitless and have warranted revisions of enormously well-confirmed and fruitful principles of logic and methodology.

2.0 The concept of presemantics in relation to proper names In 1969, Saul Kripke gave a now famous series of lectures in which he argued that there are non-analytic necessary truths. The boldfaced, indented passage ("KA) contains a summary of Kripke's arguments an outline of an argument to the effect that, given Kripke's insights, direct reference theory is *de riguer*. The latter argument not only identifies *a* reason why Kripkeans should accept direct reference theory, but *the* reason, or at least one of the more important reasons, why, despite having deep misgiving about direct reference theory and feverishly trying to make a case for an alternative, many authors, myself included, finally threw in the towel and accepted direct reference theory:

(KA) The ancients referred to the last celestial body (besides the sun) to disappear from the morning sky as 'Phosphorous' (or, to be precise, they referred to it with an expression whose Anglicization is 'Phosphorous'), and they referred to the first celestial body to appear in the evening sky

as ‘Hesperus’ (same qualification). The former is Venus and so is the latter, and the former is therefore *identical* with the latter. The ancients didn’t yet know this; and they went for a very long time having no idea that some *one* celestial body appeared in the evening sky before any other and disappeared from the morning sky after any other. But eventually they found this out. What they found out isn’t trivial, of course; it isn’t in the same category as the fact that Hesperus is identical with Hesperus. What they discovered is a non-trivial, empirical fact.

Henceforth, ‘FES’ and ‘LMS’ will be our abbreviations for the predicates ‘first celestial body to appear in the evening sky’ will be abbreviated as ‘FES’ and, respectively, ‘last celestial body (besides the sun) to disappear from the morning sky.

Contrary to what it is extremely tempting to believe,

(HIP[476]) ‘Hesperus is Phosphorous’

does *not* have the same meaning as

(S*) ‘the LMS is identical with the FES.’

It will take a moment to make it clear why this is so.

Let P be the proposition expressed by HIP. As we just saw, empirical investigation was needed to establish the truth of P; so P is an empirical truth and is therefore non-analytic. It’s important to keep the distinction between HIP and P in mind. The ancients didn’t have any feelings about the English sentence ‘Hesperus is identical with Hesperus.’ And, in making the astronomical discovery that we’ve been discussing, what they found out had nothing to do with HIP – or any other sentence. It was an

astronomical, not a linguistic, discovery. This point is crucial.

Moving on: If HIP has the same meaning as S*, then P is identical with the proposition meant by S*. And the latter is equivalent with the proposition that:

P*: exactly *one* thing is an LMS and exactly one thing is an FES, and nothing that has the one property lacks the other. [477]

And the same is true of P, supposing that S* and HIP express the same proposition.

In light of these points, consider a world W* satisfying the following conditions. W* is *semantically* just like our world. So, in W*, people speak English. But, in W*, *Venus* isn't the LMS or the FES. In W*, because of some cataclysm that didn't occur in our world, Mars is the LMS, and Mercury is the FES. So, in W*, "the LMS" picks out Mars, and "the FES" picks out Mercury. This follows from the fact that W* is semantically just like our world. (If John McCain had become president in 2008 instead of Barack Obama, 'the U.S. president' would, given the semantic rules that we use, refer to McCain. But, given those very same rules, 'Barack Obama" would still refer to Barack Obama. In other words, if x is the person who *in our world* is referred to as 'Barack Obama', 'Barack Obama' would refer to x in W*. Similarly, if Mars were the LMS, as opposed to Venus, 'the LMS' would, given the semantic rules that we use, refer to Mars. But given those same rules, 'Venus' would still refer to Venus. In other words, if x is the planet that *in our world* is referred to as 'Venus', 'Venus' would refer to x in W*.

'Venus', like 'Barack Obama', is thus a *rigid designator*. If x is what 'Venus' *in fact* refers to, x is what 'Venus' would refer to in *any* world

semantically like ours. But 'the LMS', like 'the current U.S. President', is *not* a rigid designator; it refers to different things in semantically identical situations. (It used to refer to Carter; now it refers to Obama. But the semantic rules governing that expression haven't changed in the least in the mean time.) In general, proper names are rigid designators. 'Hesperus' and 'Phosphorous' are proper names. Therefore, *unlike* 'the LMS' and 'the FES', they are rigid designators. So as long as the semantic rules of English haven't changed, there is some one object x such that (i) each refers to x; and, therefore, such that (ii) each of 「Hesperus has phi」 and 「Phosphorous has phi」 is true exactly if x has phi, a corollary being that (iii) 'Hesperus is Phosphorous' (HIP) is true exactly if x is identical with x.

It's obvious that, for any object x, the proposition *x is identical with x* is necessarily true. But it's equally obvious that the proposition expressed by HIP is empirical and, therefore, non-analytic. Therefore, HIP expresses a non-analytic *necessary* truth.

2.1 <u>Kripke's argument evaluated</u> This argument involves a massive non-sequitur, which the following story exposes. I meet a man. We shake hands. I obviously see what he looks like. He tells me that name is 'Hesperus.' (Without worrying about the details, let's assume that I know him to be telling me the truth.) In this context, there is some individual x such that I am learning that:

(HX) 'Hesperus' is x's name.

But *what* I am learning must be distinguished from the information *through which I am learning it.* The information relayed to me by my senses isn't

confined to HX. It would be absurd to think otherwise. That information is more along the lines of:

(BHX) There is some individual x such that x is standing right in front of me (in my office in Richmond, VA) at this moment (3:00 p.m., March 13, 2009) and such that x has such and such characteristics (he has a mustache, he's over 6 ft tall, etc.) and such that I am being told, by x himself, that 'Hesperus' is x's name.

So there is some x such that 'Hesperus' merely labels x. 'Hesperus' doesn't have for any part of its meaning anything having to do with the property of being over 6 ft., having a mustache, being in this or that place at this or that time. But the information *through which* I learn whom 'Hesperus' labels *does* concern such properties. That information is descriptive; a certain person x is visually represented, and thus described, to me as having certain characteristics (over 6 ft tall, etc.); and, with this help of this information, it is also being made clear to me that 'Hesperus' names x. But that descriptive information is not *itself* a part of the semantic rule that, with its help, I am learning – that is to say, it is not itself any part of HX. HX is extremely threadbare; it merely puts a label on a certain object. So the just-mentioned descriptive information has no place in the *semantics* of 'Hesperus', its relevance is confined to the presemantic act of assigning it to the right object. That information isn't semantic; it's *pre*semantic.

The following evening, I receive a phone call. I don't recognize the caller's voice. The caller is obviously using a voice-modulation device, giving it a metallic, threatening sound. And threaten is just what he does; for he says 'if you don't move out of town, I'm going to kill you. By the way, my name is Phosphorous.' I look at my caller ID; and, sure enough, the call is

from one ‘Phosphorous.’

In this context, there is some individual y such that I am learning that:

(PX) ‘Phosphorous’ is y’s name.

PX merely registers the fact that a certain name is associated with a certain object. It doesn’t say anything (else) *about* that object. It doesn’t say whether that object is tall or short, friendly or unfriendly. PX is exceedingly threadbare. But the information *through* which I am learning PX is very different from PX itself. *That* information – that *pre*semantic information - is along the lines of:

(BPX) There is some individual y such that, a moment ago (7:00 p.m., March 14, 2009), y rang me; such that, using a voice-modulation device to disguise his voice, y is making threatening statements; and such that I am being told, by y himself, that ‘Phosphorous’ is y’s name.

Here must we repeat (*mutatis mutandis*) what we said a moment ago. There is some y such that ‘Phosphorous’ merely labels y. ‘Hesperus’ doesn’t have for any part of its meaning anything having to do with the property of talking on the phone with me at a certain time, or of using a voice-modulation device, or of making threats But the information *through which* I learn who ‘Phosphorous’ labels *does* concern such properties. That information is descriptive; a certain person y is perceptually represented, and thus described, to me as having certain characteristics (calling me at a certain time, making certain statements, etc.) and, with this help of this information, it is also being made clear to me that ‘Phosphorous’ names x. But that descriptive information is not *itself* a part of the semantic rule that, with its help, I am

learning – that is to say, it is not itself any part of PX. PX is extremely threadbare; it merely puts a label on a certain object. So the just-mentioned descriptive information has no place in the *semantics* of 'Phosphorous', its relevance is confined to the *pre*semantic act of assigning it to the right object. That information is entirely *pre*semantic; not of it makes it into semantics of 'Phosphorous.'

But such *pre*semantic information can have an incalculably profound effect on what is (non-semantically, i.e. non-literally) *conveyed* by utterances. A continuation of our story makes this clear. Right after getting the upsetting phone-call, the police call. They know about the phone call (let's not worry about how), and they say to me:

(A1) 'Phosphorous is a professor of political science.'

There is some y such that A1 does nothing more than affirm the innocuous proposition that:

(A2) y is a professor of political science.

But under the circumstances, what will be conveyed to me is some much richer proposition along the lines of:

(A3) There is some individual y such that, a little while ago (at around 7:00 p.m., March 14, 2009), y rang me; such that, using a voice-modulation device to disguise his voice, y is making threatening statements; such that I am being told, by y himself, that 'Phosphorous' is y's name; and, finally, such that y is a professor of political science.

The presemantic information embedded in BPX affects what I *take away* from the policeman's utterance of L1, even though it is no part of its literal meaning. For exactly similar reasons, the presemantic information embedded in BHX affects what I would *take away* from an utterance of

(B1) 'Hesperus is an avid tennis player.'

There is some x such that B1 does nothing more than affirm the descriptively impoverished proposition that:

(B2) x is an avid tennis player.

But under the circumstances, what that utters conveys to me conveyed to me is some much richer proposition along the lines of:

(B3) There is some individual x such, a couple of days ago (March 13, 2009) at about 3:00 p.m., x is standing right in front of me (in my office in Richmond, VA; such that x has such and such characteristics (he has a mustache, he's over 6 ft tall, etc.); such that 'Hesperus' is x's name; and such that x is a political science professor.

We need only add one finishing touch to our story to see what's wrong with Kripke's reasoning. The police arrest Phosphorous. They ask me to come down to the station to talk to them. They point to Phosphorous (who is handcuffed) and say: 'that's Phosphorous.' Phosphorous, I now see, is none other than Hesperus. The man I see is the mustachioed man I met a few days earlier. The police then say: 'Phosphorous also goes by the name of Hesperus. His legal name is 'Hesperus Phosphorous.' But he tends to

introduce himself as either ‘Hesperus’ or ‘Phosphorous.’ In any case, *Hesperus is Phosphorous.*’ The italicized sentence is none other than HIP, of course.

Given what we’ve already, it obvious that there is some *one* individual z such that this utterance of HIP is true exactly if:

(K) z is identical with z.

Any proposition of that form is trivial. It is always trivial to say of an object that it is identical with itself. But by obvious extensions of what we said a moment ago, when I hear the policeman’s utterance of HIP, what I *take* away from it – what it (non-literally) *conveys* to me, given the presemantic information at my disposal – won’t be at all trivial, as it will be along the lines of:

(K2) There is some individual x such, a couple of days ago (March 13, 2009) at about 3:00 p.m., x is standing right in front of me (in my office in Richmond, VA; such that x has such and such characteristics (he has a mustache, he’s over 6 ft tall, etc.); such that ‘Hesperus’ is x’s name; and such that x is a political science professor; and there is some individual y such that, a few hours ago (at around 7:00 p.m., March 14, 2009), y rang me; such that, using a voice-modulation device to disguise his voice, y is making threatening statements; such that I was told by y himself that ‘Phosphorous’ is y’s name; and such that y is a professor of political science; moreover, I have just learned from the police that x *is* y.

K2 is an empirical proposition; it isn’t analytic; *a fortiori* it isn’t trivial. *But it’s also a contingent truth.* K, on the other hand, *is* a necessary truth. *But it’s*

also analytic. So we don't have *anything* that is both necessary *and* non-analytic.

Kripke and, after him, many others have professed to find many examples of non-analytic necessary truths; but in each case, an exact analogue of the argument just given shows a confusion similar to the one just identified to be at work. In each case, semantics is being confused with presemantics.

The Criterion of Cognitive Meaning

The logical positivist's analysis of meaningfulness was a complete failure. The problems with it are well known and needn't be repeated.[478] I'd like to propose an alternative to it.[479]

First of all, when asked to give examples of meaningless statements, the logical positivists tended to cite sentences that nobody ever uses; for example, "the nothing nothings"[480], "the all is one," "the absolute is perfect." This is deeply suspicious: a theory that only takes on straw men can't be much of a theory.

And LP can't even prevail against these straw men. For, contrary to what its advocates said, the problem with these so-called statements is *not* that they're incapable of empirical corroboration. Consider the sentence:

(i) "the universe is a perfect unity."

Much loved by many a freshman narco-intellectual, this is a meaningless sentence if ever there was one; and it is just the sort of sentence that logical positivists had in mind.

But if it were said what exactly it means to describe something as a "perfect unity," (i) *would* be meaningful, as it would then be true or false. If, by a "perfect unity", one means an object that consists of events bearing a specified causal or logical relationship to one another, then (i) is either true or false, depending on the identity of that relationship. For example, if a "perfect unity" is an object such that, given any two non-simultaneous events composing it, there is a possible causal process connecting the first of those two events with the second, then (i) is true. (In contemporary physics, ‹x precedes y› is defined as: ‹there is a possible causal process, e.g. a light-signal, beginning with x and ending with y.›) On the other hand, if, by a "perfect unity," one means an object such that, if x and y are any two of its parts, the very idea of x's existing in the absence of y is an incoherent one, then (i) is meaningful—and false. (One can coherently imagine a universe in which Cheney exists but Biden does not.)

In any case, the term "perfect unity" is clearly intended to refer to some sort of causal or logical integratedness, and once that mode of integration is pinpointed, (i) speedily becomes a true or false claim.

"Statements" such as "the universe is a perfect unity," "the nothing nothings," and so on, aren't really statements at all. They're ***statement-forms***. (A statement-form is what results when one of constants in a *bona fide* sentence is replaced with a variable. So "x is an even number is a statement-form, since it is what results when the occurrence of "4" in "4 is an even number is replaced with a variable. "4 is an even number" is an *instance* of the statement-form "x is an even number", and so is "8 is an even number." Unlike their instances, statement-forms are neither true nor false.) But (i) is obviously supposed to attribute some property to the universe. But since this property isn't identified, (i) contains an undefined term and therefore says

nothing. Once that term is defined, a meaningful statement results.

It's true that (i) is neither tautologous nor confirmable. But that's a *consequence* of the real problem, viz. that "perfect unity" is undefined. (i) is comparable to "x is tall." The reason "x is tall" says nothing is that "x" is undefined—it hasn't been assigned a referent. And no sooner is a referent assigned to "x" than "x is tall" becomes meaningful. ("x is tall" comes to have the same meaning as "Bob Dole is tall" the moment Bob Dole is assigned to "x.")

Before "x" is assigned a referent, "x is tall" is neither confirmable nor tautologous. But that's only a symptom of the real problem, viz. that "x" doesn't have a referent. The same thing *mutatis mutandis* is true of (i).

Let's move onto the next phase of our argument. If S is a meaningful sentence, there is some object x and some property phi such that S says that x has phi. In other words, any given sentence is equivalent to one that has the form: ‹x has phi.› Let us now say why this is so.

Any non-compound sentence (i.e., any sentence that doesn't consist of other sentences) either says of some individual that it has some property or it says that two or more individuals are interrelated in a certain way. Thus, "Smith is tall" says of some individual (Smith) that he has a certain property (tallness), and "Bob loves Sally" says that one individual (Bob) bears a certain relation (that of loving) with respect to some other individual (Sally).

"Smith is tall," "Jerry snores," and all other non-relational, non-compound sentences obviously have the form ‹x has phi.› And, though it isn't obvious, the same is true of "Bob loves Sally," "Wilma detests Linda," and all other non-compound sentences that affirm the existence of relations between two or more objects. Let "R" be defined as follows: for any objects x and y, ‹ <x,y> has R› is true iff x loves y. (So for any objects x and y, the ordered pair <x,y> has R exactly if x loves y.) Thus, "Bob loves Sally" is

equivalent with "<Bob, Sally> has R," which has the form ‹x has phi.›

A similar procedure can be performed on sentences (such as "Bob is standing in between Sally and Larry") that affirm the existence of relations involving three or more objects. Let "R*" be defined as follows: for any objects x, y, and z, ‹<x,y,z> has R*› is true iff x is standing in between y and z. Thus, "Bob is standing in between Sally and Larry" is equivalent with "<Bob, Sally, Larry> has R*," which has the form ‹x has phi.› Other non-compound relational sentences are to be dealt with similarly.

What about compound sentences? Not a problem. Let "K" be defined as follows: for any sentences S_1 and S_2, ‹ <S_1, S_2> has K› is true iff the state of affairs described by S_1 is a consequence (of some kind or other) of the state of affairs described by S_2. Thus, "Smith broke his leg because he fell out of the a tree" is equivalent with "<Smith broke his leg, Smith fell out of the a tree.> has K," which has the form ‹x has phi.› Other compound sentences are to be dealt with similarly.

Negative sentences are particularly easy to deal with. "Smith doesn't smoke" is equivalent with "the proposition that Smith smokes is false," which clearly has the form: ‹x has phi.› Other negative sentences are to be dealt with similarly.

There is only one kind of sentence that we haven't yet considered, namely, ***quantified generalizations.*** A "quantified generalization" is any statement that says how many members one class of objects has in common with some other class of objects. Examples are: (a) "some person smokes," (b) "no giraffes fly," and (c) "all mice read Tolstoy." (a) says that the class of people has at least one member in common with the class of smokers. (b) says that the class of giraffes has no members in common with the class of things that fly. And (c) says that the class of mice has no members in common with the class of things that don't read Tolstoy.

Bearing this in mind, let “E” be defined as follows: for any properties P and Q, ‹ <P, Q> has E› is true iff the class of things having P has a least one member with the class of things having Q. Thus, (a) is equivalent with: “<the property of being a person, the property of being a smoker>, has E” which obviously has the form ‹x has phi.› Other quantified generalizations are to be dealt with similarly.

We have thus established that any given sentence S is equivalent to some sentence having the form ‹x has phi.› Given this fact, there is an obvious answer to the question “what is it for a sentence to be meaningful?” A sentence is meaningful if it attributes some property to some object. A sentence S is meaningful if, for some object x and some property phi, S says that x has phi.[1] It’s irrelevant whether it can be perceptually confirmed, let alone verified, that x has phi.

How could this theory be wrong? If a sentence attributes any property to any thing, it says something about something and is therefore meaningful. And if a sentence doesn’t attribute any property to anything, it doesn’t say anything about anything and is therefore meaningless.

A problem with the Frege-Russell analysis of number

Frege identified numbers with sets of sets. We’ll see that he was wrong to do so and should have identified them properties of properties.

First a summary of Frege’s analysis: If S is the least inclusive set containing Jim’s shoes, then Jim has 2 shoes exactly if its membership can be put into a one-one correspondence with the membership of any two-membered set. (A “two-membered” set can be non-circularly defined as a set

S* such that for some x, y (x≠y), x and y belong to S* and such that w belongs S* iff w=x or w=y.) So we can identify the number two with the set of all two membered sets, and we can identify Jim's having two shoes with S's being a member of the number. In general, a whole number n is identical with the set of all n-membered sets. Given that statements about complex numbers, irrationals, rationals, and signed integers can be represented as statements about integers, this analysis of Frege's made it possible to represent as a set-theoretic truth any truth of arithmetic.

But there's a problem. If numbers are identified with sets rather than properties, it follows that the number two *could* have been something other than what it actually is. Consider some pair of objects – e.g. your two favorite socks – and let be the set containing just those two things. Given a universe where those two socks don't exist, if K2 is the class of all pairs in that universe, then K2 won't be identical with the class of all pairs in *our* universe. Since two sets cannot be identical unless they have the same members, it follows that *the number two could have something other than what it is*, which is absurd. Setting aside the number zero (which is the empty set and whose membership, being non-existent, doesn't vary from universe to universe) the same thing *mutatis mutandis* holds of any other whole number: if N is the class of all (and only) N-membered sets, N will be different things in different universes, since there are different things (different apples, socks, etc.), and thus different N-tuples, in different universes. But whatever the number two is in this universe, it's got to be that thing, and nothing else, in every other universe. Nothing about the number two's composition can depend on how many socks I have. So Frege's identification of numbers with sets of equipollent sets is false, since it has the false consequence that the number 5 could have had a composition different from its actual one.

In a world W where there are only eight objects, the set of all 27-

membered objects is something other than what it is here: in W, it's the empty set. In our world, of course, it is not the empty set. So the need for the Axiom of Infinity is a consequence of the fact that, as Frege analyzed them, numbers are modally elastic.

But this problem with Frege's analysis is easily remedied. Numbers are to be identified, not with sets of sets, but with properties of properties. The number two is the property of being a pair (a two-membered set). For any whole number N, N is the property of being an N-tuple. For any N, the *property* of being an N-membered set doesn't vary from universe to universe. And Frege's system, thus modified, doesn't have the absurd consequence that the composition of the number two depends on how many socks I own. Otherwise, this modification has no effect on Frege's system.

Though they evidently felt otherwise, Frege and Russell did not have the option of getting rid of properties. They had to countenance their existence. As each said many times, there is no way to identify the membership of an infinitely large set except by identifying some property common to all and only its members. And, while it's all very well to identify functions with sets of ordered pairs, there is no way to list the ordered pairs supposedly constitutive of (say) $F(x)=x^2$. It's often said that sets do, where properties do not, have clear "individuation conditions." (S and S* are the same set iff they have the same members. But how is it to be decided whether two properties are identical?

This question is ill-conceived. Mathematics deals constantly with infinitely large sets, whose members cannot be listed and can be identified only by citing some property that they have in common. So, even if properties aren't individuated with maximum precision, they are totally indispensable to mathematics. Which is not to mention that, if the number just given is correct, numbers are properties. And, further, sets, though

sometimes identified extensionally, cannot possibly be so identified for practical (or, probably, for theoretical purposes).

"But," it might be proposed, "numbers can still be identified with sets of similar sets. The number two can be identified with set of all pairs, actual *and* possible. This variant of Frege's analysis isn't open to the objection just made." If by a "possible set" is meant an *actual* set in some other (Lewisian) universe, then this amendment of Frege's analysis is in worse shape, methodologically speaking, than Frege's own analysis. On the other hand, if by a "possible set" is meant a set that doesn't exist but could, then this amendment of Frege's analysis is, like Frege's own analysis, inconsistent with the relevant modal facts. For Russell and Frege, intensions were bad and extensions were good. Properties—bad. Sets—good. Why? Many reasons. First of all, whereas it may be questioned whether there is such a thing as the *property* of being an apple, it cannot be questioned whether there is a set of all apples. Also, supposing that there is such a property, nothing is known about it---except that its what apples *qua* apples instantiate. And no content has been given to the word "instantiate." So the word "property" seems to refer to something-we-known-not-what that, if it exists, does a lot of explanatory work and that, if it doesn't, means that a lot of our explanations are defunct.

There are other, equally trenchant reasons to countenance the existence of properties and other supposedly mystical abstracta. Mathematical logic, so-called, deals with formal implication. The term "formal implication" is better identified through examples than through definitions. "If Smith is tall and fat, then Smith is fat" is formally true, since it's an instance of a form ("if x is phi and psi, then x is psi") all of whose instances are true, whereas "if Smith knows anything, then he's sentient," though analytically true, is not formally so, since it has the same form as "if Smith knows anything, then

he's a frog," which is false.

Without delving into Godel's famouis incompleteness theorems, it's not hard to show that so-called "formal (analytic) truth" is parasitic on informal analytic truth. Let T be a token of the sentence:

(*) If P, then either P or Q.[481]

Considered by itself, T is just so much ink (or noise). It's formally true *given that there are semantic rules that assign certain meanings to "or," "if," etc.* So *given* that,

(#) semantic rules being what they are, "P or Q" is compatible with not-Q,

it follows that that (*) is formally true. But (#) isn't formally true, since it has the same form as "yacht club rules being what they are, 'P or Q' is compatible with not-Q." And absent an informal truth, such as (#), (*) is meaningless and therefore not true, formally or otherwise.

So-called "formal truths," which were once thought to be identical with all analytic truths, are contingent on the existence of propositions, which are metaphysically every bit as hazy and "unacceptable" to precisionists, such as Quine, as properties.

But there are more problems with philosophical-logical-mathematical antipathy towards properties. Quine spurned propositions, saying that their individuation conditions are poor, whereas those of sentences are good. That is not good reasoning. A sentence (or sentence—token) is not a deposit of ink (by "deposit of ink" I mean any physical object that can token an expression); it's a deposit of ink that has a certain meaning and to which, therefore, semantic rules assign meaning. (Quine's false (1960) view that speech

consists of Pavlovian reflexes, unmediated by thought, insulated him from his having to realize that expressions are *ipso facto* assigned meaning by semantic rules.) So given only that two two deposits of ink have the same morphology, it no more follows that they token the same sentence than it follows that grass is Kermit the frog, given that they're both green. And since sentences are individuated by their meanings, which are given by semantic rules---which are propositions to the effect that such and such *means* thus and such---sentences' individuation conditions are *at least as bad* as those of propositions. In fact, they are quite a bit worse, given that sentences are typically replete with context-sensitive parameters that completely undermine the inference from "two sentence-tokens have the same morphology" to "they're both tokens of some one sentence."

There is an even deeper problem with the anti-property, anti-proposition conception of mathematical rigor. First of all, properties exist. If they didn't, obviously valid inferences such as

(&) given that Jeff and Jim and are conservatived, there is something (i.e. some property) that they both have in common.

No worse philosophical failures are to be found than the attempts to recover the contents of such statements *without* countenancing properties. (Hartry Field says that there are no meanings, and that languages are patterns of noises, whose "meanings" are defined by their causal roles. This view has the consequence that, when people speak, they are to no degree guided by the meanings of the words they choose. Fields view is nonetheless popular, being shared by Brandom (1994) and Soames (2004). But it's popularity is simply a blight on the discipline of philosophy. That view, as Fodor and Lepore showed in the *Compositionality Papers*, has the consequence that, if people

start saying that Aristotle wrote *Hamlet*, then the meaning of "Aristotle" so changes that anything that didn't write Hamlet isn't the philosopher of yesteryear.) Attempts to identify properties with scattered objects---e.g. Quine's attempt to identify the property of being water with the 'scattered object,' supposing that there is such a thing, consisting of all the water in the universe---makes it impossible to explain why *this* object's being water has the causal effects it does. If its being water consists in its being a member of some vast object, which is causally inert in this context, its being water is causally dead. Plus, the conditions that a thing must meet to be water don't change when the amount or distribution of water in the universe changes. So the property of being water---what it is to be water---doesn't change when that putative 'scattered object' changes, showing that that property isn't that scattered object, or anything remotely like it.

Russell (1920) said that the number 2 couldn't be identified with the property of being a pair, since there was no one property that all pairs have in common. Russell's contention is that there are many properties had in common by all pairs, but that there all pairs (two-membered sets) unambiguously are all instances of some one thing (namely, the set of all such sets). This is not good reasoning. Giiven any two pairs, there are many properties the one has that the other lacks. (One is a pair of shoes; the other is a pair of socks.) Plus, even if pairs have many properties in common other than being pairs, the property of being a pair is something that all and only pairs share, it being irrelevant that they might also have the property of (say) being Jimmy's favorite sets.

Bivalence proved, some paradoxes solved, and a fallacy refuted

1.0 Why it is a fallacy to hold that *bona fide* sentences can be meaningless

According to some, there are *bona fide* sentences that are meaningless. Among the various confusions underlying this totally erroneous position, three can be cleared up right away:

(i) There are expression-*like* entities that, given their phonetic properties along with facts about the linguistic environments in which they occur, are easily mistaken for *bona fide* expressions. Given what the words flanking it are and what it sounds like, the occurrence of "blerko" in "sometimes a blerko can be a man's best friend" is likely to be mistaken for an actual word. But it has no meaning and is thus no word.

In many lectures of his that I attended (as a student at UCLA), David Kaplan, a well known semanticist, insisted that the opening sentence of Lewis Carroll's *Jabberwocky* (" Twas brillig, and the slithy toves/Did gyre and gimble in the wabe") was a *bona fide sentence*. It is not. It is a sentence-impostor. (Obviously "brillig," "slithy," "tove," etc. are meant to bring to mind the words "brilliant," "slimy," "toad," etc. But those two lines no more token a sentence than a man wearing a Richard Nixon mask is Richard Nixon.)

(ii) According to Carnap (1932), "this stone is thinking of Vienna" is literally meaningless, even though it is a *bona fide* sentence. Carnap is making the familiar mistake of confusing the property of having a meaning worth

affirming with the property of having *no* meaning. That stone was *not* thinking of Vienna; so Carnap's sentence was false, and therefore meaningful.

There is obviously a sense in which an utterance of "this stone is thinking about Vienna" is "nonsense." But, in that context, "nonsense" means "obviously false," or the equivalent.

Chomsky (1957) made the very same mistake, though in a less brazen way, when he famously said that "colorless green ideas sleep furiously" is a sentence (since it has the right syntactic properties) but is meaningless. On one of its two disambiguations, it says that *there are ideas that are both colorless and green, all of them furious sleepers.* Thus disambiguated, "colorless green ideas sleep furiously" is meaningful, but its meaning isn't worth affirming.

On its other disambiguation, it says that anything that is a member of the class of colorless green things that are ideas is also a member of the class of things that sleep furiously. In other words, it says that nothing that falls into the first class fails to fall into the second class.

Thus disambiguated, that sentence is true but trivial: given that the first class is empty, it follows vacuously that there is no member of the first class that isn't also a member of the second class. It's obvious that the first class cannot possibly have any members and therefore cannot possibly have any members in common with any other class. So, thus disambiguated, that sentence affirms a proposition that, being vacuously true and therefore devoid of significant information, is not worth affirming.

Thus, contrary to what Chomsky says, "colorless green ideas sleep furiously" is perfectly meaningful: in fact, it's hyper-meaningful---it has two meanings. One of them is a correct (but trivial) proposition; the other is a false (and absurd) proposition. Triviality and falsity are properties of

meanings.

So Chomsky’s contention is proven false---and not, contrary to what he seems to think, proven correct---by the very fact that, depending on how it is disambiguated, “colorless green dreams sleeps furiously” is absurd or trivial.

(iii) Oftentimes, a homonym of an actual word occurs in a context in which no semantic rule assigns any meaning to it and in which it therefore means nothing, even though the context in question is one that is *otherwise* just like a sentence. What occurs in “the absolute is one” is not the word “absolute,” but a mere homonym of that word. For the occurrence of “absolute” in “the absolute is one” to be a word, there would have to be some semantic rule that assigned it a meaning. Although there are semantic rules that assign meaning(s) to homonyms of that occurrence, that occurrence *per se* is assigned no meaning. Anything that is assigned no meaning is *ipso facto* not a linguistic expression. So “the absolute is one” is not a sentence, but a sentence-impostor, at least until it is said what "the absolute" refers to, or is supposed to refer to---at least, that is to say, until the semantic rule for it is identified.

For---and this cannot be stressed enough---anything to which no semantic rule assigns a meaning *ipso facto* isn't a linguistic expression. Because the logical positivists overlooked this humdrum fact, they sought some *criterion* by which to distinguish "meaningful sentences" (e.g., "the moon is made of cheese") from "meaningless sentences" (e.g., "the absolute is one"). But all of the meaningless so-called sentences that they identified weren't sentences at all; they were sentence-schemata, since they contained undetermined expressions (expressions to which no semantic rule assigned any meaning, at least not as far as occurrences in the sentence in question were concerned) and therefore no more qualified as *bona fide* sentences than does "*x* has

property phi."

2.0 The liar's paradox

We've found elementary blunders in each of the reasons already considered for holding that *bona fide* expressions can be meaningless. But there are other reasons for holding this that don't involve any obvious blunders and demand careful treatment. In the process of considering these reasons, we'll make some headway on some important, open questions in semantics and philosophical logic.

There are sentences that, so initial appearances suggest, are true if they're false and false if they're true.[482] Consider the sentence:

(LIE_1) "I am now lying."

Given the law of excluded middle ("any proposition must be either true *or* false"), LIE_1 cannot fail to be either true or false. Given the law of non-contradiction ("no proposition can be both true *and* false"), LIE_1 cannot be both true *and* false. So if LIE_1 expresses a proposition, that proposition is both true and false. Therefore, given that no such proposition exists, LIE_1 doesn't express a proposition. But it's very hard to believe that "I am lying" is not a sentence. Exactly similar reasoning holds with respect to "I am lying" and "I am a Cretan and all Cretans are liars," and each one of the infinitely many other sentences that are relevantly similar to those two.

LIE_1 is indeed a sentence. But it's not both true and false; it's just false. LIE_1 is equivalent with:

(LIE_2) “There already exists some proposition *P* such that *P* is false and such that, by virtue of producing this very utterance, I am affirming *P*.”

LIE_2 thus entails that, existing prior to, and therefore independently of, any LIE_2 or any utterance thereof, there is some proposition *P* such that (i) is false and, moreover, such that (ii) is true only if, for some utterance of LIE_2, *P* is true only if that utterance is true. Since *P pre*-exists any such utterance, LIE_2 is not false *and* true; it is simply false. Therefore, LIE_1 is false. So we don’t have in LIE_1 a counter-example to our thesis that there are *bona fide* expressions that are meaningless.

2.1 The liar’s paradox (continued)

There is another way of dealing with the Liar’s paradox, or at least with one version of it. Let UTF be a specific utterance of the sentence:

(TF) “This sentence-token is false.”

To what do the first two word-tokens constituting UTF refer? “They refer to UTF,” it will be said. But UTF doesn’t yet exist. Suppose that somebody---call him “Smith”----whose intent on tokening is interrupted immediately after saying “this sentence-token,” and thus cannot token TF. (So, having said “this sentence-token,” Smith does not say “is false,” at least not within the appropriate time-period.) To what does this occurrence of “this sentence-token” refer? Nothing.

“Maybe,” it will be said, “it refers to the sentence-token that he *would* have tokened.” In that case, he’s still referring to nothing and has therefore said nothing. For, supposing that he’s referring to an entity that *would* have come

into existence but didn't, there *ipso facto* does not exist an entity to which he *is* referring and, as previously stated, he is thus referring to nothing. Given that he's referring to nothing, what he's saying no more expresses a proposition than does a token (that isn't preceded by a token of "this sentence-token") of "is false." In other words, Jones has, for all intents and purposes, said "is false"---nothing more. Thus, he hasn't affirmed anything or, therefore, anything self-contradictory.

"Doesn't that show that Smith's utterance is meaningless?" it will be asked. "And, given that his utterance is obviously a *bona fide* expression, doesn't it follow that, contrary to what you are arguing, there *do* exist *bona fide* expressions that are meaningless?"

Smith's utterance fails to affirm any proposition; but it isn't meaningless. If I point to an empty space and say "that elephant is big," I've failed to affirm a proposition---but not because I'd produced a meaningless utterance. There exist rules to the effect that I've affirmed a true proposition *if* certain conditions are met. So that token has certain truth-conditions and is *ipso facto* meaningful. A *consequence* of the fact that my utterance has those truth-conditions is that, given the circumstances, it fails to affirm a proposition, another consequence being that, had circumstances been different, it would have affirmed a proposition. So a sentence-token's affirming a proposition is a consequence, not the essence, of its being meaningful. What one affirms in tokening a proposition-bearing sentence-token is the proposition that, given that circumstances, that token's meaning assigns to it; what one affirms is not that meaning itself.

Let us move on by considering Smith's scenario with Brown's. Brown is intent on tokening TF and, unlike Smith, isn't interrupted and thus carries out his intention. (So, right after tokening "this sentence-token," Brown tokens "is false.") Brown's token of "this sentence-token" co-refers with Smith's

token of “this sentence-token.” It is not an option to say that, because Brown manages to complete his utterance, his (Brown’s) utterance of “this sentence” has a referent (namely, his utterance of TF) that Smith’s does not. For the meaning of a sentence-token is a function of the meanings of its parts. That means that any given part of a sentence-token must have a meaning independently of what that sentence-token itself means. This, in its turn, means that Brown’s token of “this sentence-token” must have a referent independently of the meaning (if any there be) of the sentence-token as a whole in which it is embedded. And this, in *its turn*, means that Brown’s token of “this sentence-token” must co-refer with Smith’s token of that expression; which means that it refers to nothing. Brown’s utterance is, as a whole, quite as meaningless as Smith’s. (That is, Brown’s utterance is meaningless *in the sense* that there is no proposition it affirms. It is not meaningless *tout court.*) No meaning. No paradox.

3.0 Russell’s paradox

Other sentences that are obviously comparable to “I am now lying” are “this sentence is false,” “I’m a Cretan, and all Cretans are liars.” But there exist sentences that, it has been argued, are meaningless and that, unlike those two, are not in any obvious way modeled on “this sentence is false.” The best known example of such a sentence is (‘RP’ stands for “Russell’s Paradox”):

(RP) *K* is a member of itself iff *K* is not a member of itself,

where *K* is the class of all classes that are not members of themselves. (If *K* is a member of itself, then, since *K* includes only classes that aren’t members of

themselves, then *K* is not a member of itself; and if *K* is not a member of itself, then, since *K* includes every class that is not a member of itself, *K* is a member of itself.)[483]

According to Graham Priest, a contemporary logician, RP is meaningful; i.e., there is some proposition that it affirms; but, Priest says, since that proposition is both true *and* false, we must reject the law of non-contradiction.[484] In a moment, we'll see why Priest's reasoning is spurious.

Unlike Priest, Bertrand Russell held that the law of non-contradiction is correct and, therefore, that there is no proposition that RP expresses. But Russell has a very different view from us as to *why* RP fails.

Russell's view is easily shown to be false and, moreover, to be self-undermining. Russell's view is embodied in his famous "Theory of Types" ("TT" for short). According to TT, any given sentence belongs to a certain "type"; given any two types, one is higher than the other. The lowest type is Type_1. Sentences of Type_1 concern "individuals." (An individual is anything that is anything spatiotemporal and therefore anything that isn't a set, property, sentence, or proposition. In this context, for the sake of expediency, references to "classes" will be short-hand for references to classes *and* to properties *and* to propositions *and* to sentences.) Sentences of Type_2 concern individuals *and* classes whose members are individuals. (Sentences of Type_2 thus concern *inter alia* sentences of Type_1.) Sentences of Type_3 concern everything with which sentences of Type_2 are concerned *as well as* classes of classes of individuals.[485] In general, for any whole number *n*, sentences of Type_{n+1} may concern everything with which sentences of Type_n are concerned and, *unlike* sentences of Type_n, sentences of Type_{n+1} may concern classes of Type_n.

A consequence of TT is that, if the free variable in "*x* is a member of class C," where C is of Type_m, is replaced with an expression denoting a class of Type_m or higher, the resulting sentence is meaningless; i.e., there is no proposition that it affirms and it is therefore neither true nor false.

So, according to TT, "*C* is a member of *C*" is meaningless, and so is its negation. TT entails that RP is no more meaningful than "blurka derka derk." Given his acceptance of TT, Russell understandably holds it to be a failing of the English language that it permits the construction of sentences like RP.

3.1 Why the theory of types is false

RP is ambiguous; there are two ways of disambiguating it. RP can be taken to be to the same effect as:

(RP_1) There exists some class *K* such that all and only those classes that are not members of themselves belong to *K*.

Thus disambiguated, RP is false. There is no such class. A contradiction follows from the supposition that there is a class whose membership includes all and only those classes that aren't members of themselves. Therefore, there is no such class.

The fact that no such class exists is to be established by a garden-variety indirect proof, just like the fact that there are no square circles.[486] Let *x* be a square circle. Being a square, *x* has four sides and therefore doesn't have a uniform curvature. Being a circle, *x* has a uniform curvature. Therefore, *x* both does and does not have a uniform curvature. This isn't possible. So *x* is a non-entity. *K* is a non-entity for the same reason *mutatis mutandis.*

There is another way to interpret RP. We could take "*K*" to be defined

thus:

(AI[487]) *Assuming* that there is some class *O* whose members include all and only those classes that are not members of themselves, then, in virtue of having the form ‘ …*K*…’, a given sentence *S* is true just in case *O* has…*x*… (i.e., just in case…*O*…); and, if that assumption is false, then “*K*” picks out nothing, and ‘…*K*….’ bears no proposition and is thus neither true nor false.

Thus taken, RP doesn’t say *anything*. Since there is no class of the requisite kind, “*K*” picks out nothing and, the semantic rule for “*K*” being what it is, RP fails to encode any proposition. Thus disambiguated, KP is like an utterance of:

(TPE) “That pink elephant is coming right at us,”

made in a context where there is no pink elephant. So, thus disambiguated, RP requires neither an abandonment of the laws of classical logic nor an acceptance of TT.

3.2 Two kinds of meaning: Type-meaning and token-meaning (revisited)

Here we must revisit an issue discussed in the previous chapter.[488] There seems to be *some* sense in which an utterance of TPE that is made in an elephant-free context is meaningful, notwithstanding that such an utterance bears no proposition. The sense in which such an utterance is

meaningful is this: the semantic rules of the language to which it belongs determine what *would* have to be the case for that utterance to be meaningful. The semantic rule corresponding to TPE is this:

(TPE*) If, in context *C*, there is some pink elephant *E* that is being indicated by some person P_1, and P_1 is indicating no pink elephant other than *E*, then an utterance on P_1's part, directed towards persons $P_2 \ldots P_n$ ($2 \leq n$), of "that pink elephant is heading towards us" is true exactly if *E* is heading towards $P_1 \ldots P_n$. If there is no such elephant, that utterance fails to encode any proposition.

So, when uttered in an elephant-free environment, TPE *is* meaningful. It's meaningful in the sense that it has well-defined truth-*conditions.* Whether those conditions are fulfilled or not has no bearing on whether TPE is meaningful or not. The semantic rules of English don't "know," as it were, whether an elephant is present in the context in question. One could know the semantic rules of English backwards and forwards and not have the faintest idea as to whether an elephant was present in the context question. So, the semantic rules of English don't "know," as it were, whether an elephant is present in the context in question.

Similar reasoning holds in connection with RP. RP has truth-conditions. Whether those conditions are fulfilled is irrelevant to whether RP is meaningful. "But there is a difference between RP and TPE," it will be said. "Whereas tokens of TPE *could* be true, tokens of TT cannot possibly be true: the laws of logic prohibit it." Yes---but the laws of logic aren't coded into the semantic rules of English. One can master those rules without being a master-logician; one certainly doesn't need to know whether there exists a class of all classes that aren't members of themselves. So RP, though not

proposition-bearing, *is* meaningful.

To conclude: on neither of its disambiguations is RP a counterexample to our contention that all *bona fide* expressions are meaningful. This is another instance of a philosopher failing to distinguish absurdity from meaninglessness. We discussed other such examples in Chapter 1.

3.3 The theory of types revisited: More problems with it

The Theory of Types (TT) says that any meaningful sentence must occupy a higher order than any sentence that it concerns. But there is no way to state TT without breaking this very rule. For, according to TT, *any* given sentence S is meaningful only if it concerns nothing of an order equal to or greater than its own. TT thus falls within its own scope and, supposing it correct, is therefore meaningless (and therefore *not* correct) for the same reason as RP. Nothing meaningless is correct. S, there is no way that TT can be correct, given that it entails that it is itself meaningless.

Also, TT has consequences too counterintuitive to accept. For example, according to TT, the sentence:

(i) "Jim has three pairs of socks."

is of a lower order than the sentence:

(ii) "The sentence 'Jim has three pairs of socks' is true,"

and (ii) is of a lower type than the sentence:

(iii) "The sentence 'the sentence "Jim has three pairs of socks" is true' is

true."

If TT is right, the property that one ascribes to (iii) by describing it as "true" is one that cannot possibly be had by (ii) or (i), and the property one ascribes to (ii) by describing *it* as "true" is one that cannot possibly be had by (i). So, if TT is right, there is no such thing as truth *simpliciter*. There is truth_1, truth_2, etc.

By similar reasoning, if TT is right, the occurrence of "three" in:

(iv) "Jim has three pairs of socks,"

doesn't have the same meaning as its homonym in (i). (i) is to the effect that:

(a) There are three true instances of the propositional-form *x is a member of the class of pairs of socks had by Jim*.

The occurrence "three" in (i) denotes a property had by classes of individuals. (It denotes the property of being a three-membered class of individuals.) The homonymous occurrence in (iv) denotes a property of classes of classes of individuals. (It denotes the property of being a three-membered class of classes of individuals.) So the property denoted by the one occurrence is distinct from the property denoted by the other. Given how counterintuitive this is, TT must be rejected.

4.0 Vagueness and indeterminacy

The word "indeterminate" has two meanings. The sentence:

(BW) “Bill’s whereabouts are indeterminate,”

can be taken to mean either:

(BW_1) given the available information, it isn’t possible to arrive at a principled judgment as to where Bill is.

or,

(BW_2) there is no fact as to where Bill is.

BW_2 cannot possibly be true. We will soon see why. At any given point in time, some space-time region *R* such that Bill is occupying *R*. Of course, half of Bill's body can be in the garage while the other half is in the drive way. But that’s because the region he is occupying is itself half in the garage and half in the driveway; it is not because there is no region that Bill is determinately occupying. The same points *mutatis mutandis* hold of any other proposition to the effect that the world *per se* is indeterminate in any respect. In any case, this is what the present author believes.

But many disagree; and their reasons for doing so seem, at least at first, to be good ones. For there certainly *seem* to be cases of objects with vague boundaries. In fact, all spatiotemporal objects, it would seem, have vague boundaries. Given any person *P*, there are many particles *x* such that one cannot non-arbitrarily say that ‹*x* is a part *P*'s body› is true *or* that it's not true. This apparent fact is one that has led many a philosopher to hold that there is objective indeterminacy.[489] But we'll now see that such philosophers didn't think deeply enough about this so-called fact.

5.1 Non-epistemic indeterminacy

Smith doesn't have a full head of hair; but there are a few tufts of hair on his head, in between which there is fuzz of varying degrees of thickness and density. Bearing this in mind, consider the statement:

(SB) "Smith is bald."

Is SB true or false? Many philosophers, probably most, would hold that:

(OV)[490] there is no fact of the matter; and that fact has nothing to do with ignorance on our part. Here's what's going on. There exists a certain property *P* that of baldness and a certain person *x*. (*P* = the property of baldness; *x* = Smith.) And *x* doesn't exactly instantiate P, but *x* doesn't exactly fail to instantiate P. There is no fact as to whether the proposition *x has P* is true. It isn't that we don't know whether it's true; it's that there is no fact of the matter.

Here's the problem with OV. For argument's sake, suppose that it is objectively indeterminate whether Smith is bald. Suppose, in other words, that the proposition (not the sentence):[491]

(SB) Smith is bald,
is true.

And, let us also suppose, the reason SB is vague is that it is objectively indeterminate exactly what conditions a given thing must satisfy to be bald.

Given these suppositions, it is *ipso facto* indeterminate what exactly SB entails and what is entailed by it. If it is objectively indeterminate exactly what features a thing must have to be bald, then it is for that very reason indeterminate, for any *x*, exactly what entails, and also what is entailed by, the proposition (not—*nota bene*—the sentence):

(XB) ‹x is bald›,

for any *x*.

Thus, a consequence of the supposition that SB is *objectively* indeterminate is that it is indeterminate what SB entails and also what is entailed by it.

But the very idea of such a proposition is not a coherent one. Propositions are individuated by what they entail. Different entailment relations, different propositions. So until the entailment relations of some sentence-like string of verbiage have been fixed, no proposition has been identified to that verbiage; and all we have is, at most, a sentence-schema.

5.2 An objection to the argument just presented

I have heard it said that:

(ER)[492] Propositions are *not* individuated by their entailment relations. The proposition that 1 + 1 = 2 obviously isn't identical with the proposition that triangles have three sides. But a given proposition entails/is entailed by the one just in case it entails/is entailed by the other."

Obviously *1 + 1 = 2* and *triangles have three sides* are different

propositions. But contrary to what ER assumes, they *don't* coincide in respect of their entailment relations. The *way* [493] in which *for some n, n + 2* follows from *1 + 1 = 2* differs—quite dramatically—from the way (or, rather, any one of the infinitely many possible ways) in which *for some n, n + 2* follows from *triangles have three sides.* A corollary is that, until it is settled what the entailment relations are of a given sentence, it *ipso facto* hasn't been settled which proposition is meant by that sentence. (This doesn't mean that one has to know what all of those relations are. No one knows what *all* of the entailment relations are of "1 + 1 = 2." But, the semantic rules of the English language being what they are, it is indeed fixed what those entailment relations are.)

What a sentence says---or, to be precise, what is affirmed by a non-abortive sentence-token in the indicative mood---is a function of exactly two things: first, what follows from that sentence-token (supposing it is true); second, *how* those things follow from it (same qualification). Given a sentence-token S, once what it entails and how it entails it is fixed, there is nothing left to be said about what it means. By the same token, if what S entails or how it entails it isn't fixed, it has yet to be fixed what S means. Thus, a sentence such that there is no fact as to what it entails, or as to how it entails what it entails, is a sentence to which no one proposition has been assigned.

Bearing these points in mind, suppose the following. *P* is a true proposition. There are propositions $P_1 \ldots P_n$ such that, for each i, it is objectively indeterminate whether *P* entails P_i. Thus, *P ipso facto* leaves it open whether *P* is true. But if *P* leaves it open whether P_i is true, then *P* determinately *fails* to entail P_i. That doesn't mean that *P* entails the *negation* of P_i. (*Snow is white* doesn't entail *grass is green* or *grass is not green*. A proposition may be compatible with another proposition *and* with its

negation.)

The believer in indeterminacy has but one way of dealing with this argument. He is forced to say that, although there is no fact as to whether *P* entails P_i, nonetheless *P* does not leave it *open* whether P_i is true. But this position is tautologously false.

5.2.1 An objection to the argument just presented (continued)

Here is a similar argument. Suppose there to be some propositions *P** such that (i), for any given proposition, it *is* determinate whether *P** entails that proposition and therefore that it *is* determinate, for any *i*, whether *P** entails P_i but such that (ii), with that qualification, *P** entails a given proposition iff *P* also entails it. Is *P* identical with *P**? It cannot be. For *P** has, whereas *P* lacks, the property of being such that there is a fact as to whether it entails P_i, for some *i*. Thus, *P* is distinct from any proposition *P*% such that *P*%, supposing it true, settles whether or not P_i is true. So *P fails* to settle whether or not P_i is true and thus fails to entail it or its negation. So the statement "*P* entails P_i" is determinately false.

Were there indeterminate statements; i.e., statements such that there was no fact as to whether what they affirmed was true, there would be pairs of statements such that there was no fact as to whether the one entailed the other (or, therefore, whether it entailed its negation). But given any such pair, the one *ipso facto* leaves it open whether the other is true and thus determinately *fails* to entail (the negation of the other), an immediate consequence being that there *is* a fact as to whether the one entails the other. So the concept of a proposition such that there is no fact as to whether it is true is an incoherent one and the same is therefore true of the concept of objective, as opposed to epistemic vagueness.

6.0 The Sorites paradox

Consider the following paradox.

Paradox #1 (P_1): A person with only one dollar is poor. If a person with n dollars is poor, then so is a person with $n + 1$ dollars. It follows that a person with a billion dollars is poor. But, obviously, a person with a billion dollars is not poor.

For reasons that are of no philosophical importance and that, in any case, will soon become clear, P_1 is an example of what is known as the "Sorites Paradox" (a.k.a., the "Paradox of the Heap," "Sorites" being the plural of the Greek word "soros," which, so I am told, means "heap").

Here is an outline of the solution to P_1: Like most predicates, "poor" is a relational expression. Smith may be poor compared to Jones but wealthy compared to Green, or he may be poor *relative* to the standard set by some situation and wealthy *relative* to the standard set by some other situation. If Smith is a millionaire who needs to have a net worth of at least 50 million dollars to become a managing director of company *X*, he is poor relative to the standard of wealth that, given Smith's desire to become a managing director of *X*, is the relevant one. So an utterance of 'Smith is poor' is correct if made in a context in which the relevant standard of wealth is set by Smith's desire to become a managing director of *X*. For, in such a context, a 'poor' person is one whose net worth is less than $50,000,000. For that very reason, it follows that, in such a context, *there is* a number *n*-- namely, 49,999,999---

such that an utterance of ‹*x* is poor› is true of *x* has n dollars and false if *x* has *n* + 1 dollars.

Bearing these points in mind, it's clear how to solve each member of the infinitely large class of paradoxes to which P_1 belongs. Consider the following paradox:

(P_1#) If a person's top running speed is 1 mph, then that person is slow. For any number, if a person whose top running speed is *n* mph is slow, then a person whose top running speed is *n* + 1 mph is also slow. Therefore, a person whose top running speed is 75 mph is slow.

A story, structurally similar to the one just told, will make it clear how to solve this paradox. At his best, Smith runs a 100 yard dash in 12 seconds. Smith tries to join track team *X*. In order to qualify for team *X*, one must run a 100 yard dash in no more than 10 seconds. Obviously Smith doesn't make the team. Smith then decides to join team *Y*. In order to qualify for team *Y*, one must run a 100 yard dash in no more than 100 seconds. Smith obviously qualifies for *Y*.

If made in a context in which the relative standard of swiftness is set by Smith's desire to join *X*, an utterance of "Smith is fast" is false. But such an utterance is *true* if made in a context in which the relative standard of swiftness is set by Smith's desire to join *Y*. In a context of either kind, there is some number *n* such that an utterance of ‹*x* is slow› is true if, at his best, *x*'s average speed when running the 100 yard dash is n and such that an utterance of "*x* is fast" is true if *x* is person whose average running speed when running 100 yards is *n* + 1.

In a context of the first kind, *n* is any number such that if a person whose average speed when running a 100 yard dash is n and who, for that reason,

runs the 100 yard dash in over 11 seconds but who, if his speed when running the 100 yard dash were *n* + 1, would run it in under 11 seconds. The same thing *mutatis mutandis* holds of contexts of the second kind.

6.1 These points generalized

To generalize these points, we must note a fact about language—more specifically, about predicates.[494] Some predicates are binary. *P* is a binary predicate if the property it expresses is one that cannot be had to varying degrees; it is one that a given thing either has or doesn't have. So "even" (as in "even number") is a binary predicate. Sorites paradoxes always involve non-binary predicates (i.e., predicates that express properties that can be had to varying degrees), e.g., "fat," "tall," "smart." A corollary is that Sorites paradoxes always involve properties that one thing can have to a greater extent than some other thing. One thing can be smarter than another.

Bearing these points in mind, consider the following three sentences:

(i) Smith is smart.
(ii) Smith is *more* smart than Jones.
(iii) Smith is smart*er* than Jones.

(ii) and (iii) are synonyms. In effect, (iii) is just an abbreviation of (ii). Surely, the occurrence of "smart" in (i) has the same meaning (picks out the same property) as the homonymous expression in (ii). This wouldn't be possible if, in (i), the occurrence of "smart" picked out a binary (non-relational) property. This shows that "smart" is *always* relational. This, in its turn, shows that tokens of (i) are elliptical for sentence-tokens that express-relational propositions---sentence-tokens having the form:

(i#) “Smith is smarter than___,”

or perhaps:

(i*) “Smith is smart *enough* for__.”

So tokens of “Smith is smart” must be elliptical for sentences such as “Smith is smarter than Jones” and “Smith is smart enough to solve the puzzle.” Exactly similar points hold of other non-binary predicates. The meaning of “*x* is fast” is given by some sentence having the form “*x* is fast*er* than *y*” (e.g., “Smith is faster than Jones”) or having the form “*x* is fast *enough* to satisfy standard S” (e.g., “that plane is fast enough to make it to Pittsburgh by 3:00 P.M.”)

6.2 A confusion side-stepped

Here I would like to head off a source of serious confusion. I am not advocating, and am inclined to reject, the view that *at the level of semantics* “Smith is smart,” “Jones is tall,” etc. express relational propositions of the just-mentioned kind. It cannot be assumed that the philosophical insights needed to solve puzzles that turn on facts about language are themselves encoded in the semantic rules of English (or of whatever language happens to be the relevant one). So the just-stated analysis is not a *semantic* analysis; it is not an attempt to identify the semantic rules that assign meaning to sentences of the form “*x* is *P*,” where *P* is an arbitrary non-binary predicate.

It makes no sense to say that some number is *more* even than some other number. But this has nothing to do with language; it has to do with the

property of being even. The sentence “6 is *more even* than 8,” though absurd, does not run afoul of the semantic (or syntactic) rules of English. (Or, if it does, it is only because we have *ex post facto* projected our mathematical insights into syntactico-semantic rules of English. And were some mathematician to show that evenness came in degrees, it would come to be grammatical---if it isn’t already---to say ‹*x* is more even than *y*.›) Clearly, “two is even” is not semantically or otherwise elliptical for some sentences of the form ‹*x* is more even than *y*› or ‹*x* is even enough to satisfy standard *S*.› But unless we implausibly suppose that our knowledge that evenness doesn’t come in degrees is coded into the semantic rules of English---unless we implausibly endow the semantic rules of English, which can be mastered by somebody who wrongly believes that evenness *does* come in degrees, with our mathematical knowledge---then ‹*x* is even› is not, at the level of semantics, elliptical for “*x* is more even than *y*” (etc.).

This reasoning *mutatis mutandis* shows that *at the level of semantics* ‹*x* is fast› probably isn’t elliptical for ‹*x* is faster than *y*› (etc.). For, were we to say otherwise, we’d be endowing the semantic rules of English with our (extra-linguistic) knowledge of the fact that, indeed, swiftness comes in degrees. Given that those rules can, at least in principle, be mastered by somebody who has the idiotic belief that swiftness does not come in degrees, we mustn’t impute to the semantic rules of English our knowledge that swiftness *does* come in degrees; and we must therefore take the position, for which there is a vast amount of independent corroboration, that this knowledge of ours is non-linguistic.

That said, the way to solve P_1 is to take the view that *in some way or other*---probably through pragmatic implicature, not necessarily through literal meaning---tokens of (e.g.) “Bob is smart” are elliptical for tokens of relational sentences of the previously described kind.

6.3 A different kind of Sorites paradox

Let us now move onto a different kind of Sorites paradox. Paradoxes of this kind are much harder to solve than those of the kind just discussed:

P_2: Let *C* be some cloud and let *m* be some water-molecule dead in center of *C*. Obviously *m* is a part of *C*. But let's suppose that m migrates towards the periphery of *C*. For a while, m will still remain a part of *C*. But if *m* travels, say, two miles away from *C*, then *m* will definitely cease to be a part of *C*. But where's the cut-off line. Suppose that, for a while, *m* whizzes around *C*'s periphery, sometimes bumping into molecules that are definitely a part of *C*, and never straying from *C*'s periphery---i.e., from what *definitely* either coincides with or falls within *C*'s periphery----by more than an inch. Under those circumstances, is *m* a part of *C*? What about if we continue to suppose that *m* intermittently bumps into *C* but we increase this margin from one inch to five inches? Is *m still* a part of *C*? And how infrequently can *m* collide with molecules that are clearly *C*-constitutive without itself definitely ceasing to be a part of *C*? If such collisions are frequent, we're inclined to consider it a part of *C*, even during those periods when it's not in contact with molecules that are clearly *C*-constitutive (provided that, during such periods, it doesn't veer off by *too* great a distance). By the same token, if such collisions are infrequent, we're inclined *not* to consider *m* a constituent of *C*, even if, in between such collisions, it remains fairly close to *C*. (The more frequent the collisions, the more *m* can travel away from *C without* ceasing to be a constituent of *C*.)

But where is the cut-off line? Or, rather, where are the cut-off *lines*? (For, whether or not *m* is a member of *C* seems to depend *both* on its average

proximity *and* on the frequency with which it comes into contact with *bona fide C*-constituents.) But, for simplicity's sake, let's just focus on the first question. Where, I again ask, is the cut-off line? Is it half an inch? Two inches? A foot? What is it? Any answer seems arbitrary. If I say: "it's exactly three inches," I'm making a stipulation; I'm not identifying a pre-existing fact.

So here, quite indisputably, we have a case of indeterminacy. But, more interestingly, this indeterminacy has nothing to do with the representations of the world that we, or creatures of any other kind, have of the world. There could (and probably would) be clouds even if there were no sentient beings; and, in the absence of such beings, there would, given just about any cloud *C*, be molecules that were definitely parts of *C* and others that definitely weren't parts of *C and*, what is relevant, molecules *x* such that *x is a constituent* of *C* wouldn't exactly be true and wouldn't exactly be false---or, to make an equivalent point, such that *x's constituency in C* would not itself exactly be a constituent of the universe and also wouldn't exactly fail to be a constituent of the universe.

6.3.1 A correct conception of thinghood the key to solving P_2

The solution to P_2 is similar to the solution to P_1. But there are some significant differences; and to solve P_2, we must first make some substantive claims as to what is meant by the words such as "thing" and "object." (In this context, those terms are to refer only entities such as vases, rocks, windows---not to *abstracta* such as numbers or possibilities.) For the reasons given in Sections 5.2-5.2.1 and now to be developed, I don't consider it an option to say that the world *per se* is indeterminate. Whatever the word "cloud" means, we gave it that meaning. The corresponding semantic rule cannot possibly

contain information about clouds that we didn't build into it. That rule cannot possibly contain the information needed to substantive questions concerning cloud-composition. Open questions in meteorology can't be answered by studying semantic rules. Semantic rules cannot know more than we do. So it's understandable, if not inevitable, that there should be arrangements of molecules that determinately don't deserve to be characterized either as "clouds" or as "non-clouds," and it's understandable, if not inevitable, the word "cloud" should fail to pick out *any* one property. That word cannot our semantic conventions to cut nature at the joints.{au: is something not quite right about prev sentence? and the next?} We cannot our linguistic conventions to be hewed to exactly those properties in terms of which the constitution of the world is to be delineated. Laws created 200 years ago should fail to address issues relating to the disposal of motor-oil. There's nothing mysterious about this, and it's obviously no reason for saying that the world is intrinsically vague. The same holds of semantic rules that were created over 200 years ago. And semantic rules, like laws, must be updated if they are to do what we want them do. So it isn't hard to make sense of the idea that semantic rules should sometimes be silent.

But it *is* hard to make sense of the idea that properties should be indeterminate. What does it mean to say that *property* of being a cloud is indeterminate? Which property is the one being so described? "The one had by those billowy things in the sky." Yes, but there are other objects that this person is not currently indicating. So by what criterion do we decide, when confronted with some new object, whether it is of the same ilk as those billowy things? Even if we could see the microstructural properties of those objects, we wouldn't necessarily know in virtue of which facts about their microstructures they fell into the same category as those billowy things. Until we can answer these questions, we don't have an answer to our original

question, namely: *which* property is the relevant one? The information we have limits what a viable answer to that question could be. But it still leaves us with an infinitely large class of distinct properties to choose from.

To sum up: attempts to identify properties that are objectively indeterminate can be interpreted as cases where *no* one property has been identified, suggesting that so-called cases of objective indeterminacy are nothing more than projections into the external world of the platitude that open scientific questions are not to be solved by studying semantic rules.

6.3.2 A correct conception of thinghood is the key to solving P_2 (continued)

In any case, P_2 arises because we think about things in the wrong way. We think of things as being, to put it paradoxically, more thing-like than they really are. Right now I'm looking at the pen on my desk. Let *x* be one half of my pen (the segment whose length is half of that of pen which includes the pen's pointy tip), and let *y* be some region of the table immediately adjacent to *x*. Finally, let *Z* be the 'entity', for lack of a better word, whose only two parts are *x* and *y*. *Z* is not a 'thing.' If we were to list all the things in the world, my pen would be on the list, and so would the table on which it rests. But *Z* wouldn't be on the list. Why not? Because *Z* lacks the dynamic cohesiveness, and concomitant predictive and explanatory usefulness, that is had by the pen or the table or any other *bona fide* object.

Of course, tables don't last forever. They can be broken and burnt; and they all come to an end, in one way or another. But it is never *by virtue of* being a table that a thing ceases to exist. *Qua* table, a thing can be depended on and its behavior predicted. The fact that poorly constructed tables are not thus reliable only confirms this point; for, in virtue of being poorly

constructed, and for that reason lacking causal and explanatory integrity, a thing *fails* as a table. This only confirms that that it is never *as a table* that a thing lacks causal or predictive usefulness.

These points *mutatis mutandis* hold of *all* objects. Things are *ipso facto* prediction- and explanation facilitators. One thing can be more prediction- and explanation-conducive than another. Because of its relatively high degree of causal integrity, a house made of bricks and mortar is more explanation- and prediction-conducive than a house of cards. We are independently inclined to say that a brick house is more of a thing---is more "thingy"---than a house of cards. We now know what it is that we're saying when we say this. We're saying that a brick house (supposing it is properly constructed, of course) is, other things being equal, more cohesive, more of a unit, dynamically speaking, and—for these reasons—more durable than a house of cards. Thus, our analysis that things are prediction-/explanation-facilitators is consistent with our pre-theoretic intuition that (to switch examples) clouds, though things in some sense of the word, are less thingy than houses.

6.3.3 A consequence of our analysis

One consequence of our analysis is that a given aggregate of molecules may be a thing in one respect and fail to be a thing in some other respect. In a context where the interactions holding among molecules $m_1 \ldots .. m_n$ facilitate prediction and explanation, $m_1 \ldots . m_n$ constitute a thing. In a different explanatory context, one where they don't have the aforementioned virtues, they don't constitute a thing. A galaxy is a vast object. But a galaxy does, in some contexts, function as a single object; it has the requisite causal integrity and predictive fertility that, when trying to answer certain questions, one does well to consider the various microparticles of which a galaxy consists as

constituting a single object. In other contexts, one is ill-advised to do this, for the same reason *mutatis mutandis.*

To be sure, we believe that some aggregates of molecules (e.g., vases, rocks, pens) are *just* things; we believe, in other words, that their having the status of thing isn't relative to the dialectical context. This viewpoint reflects prejudices that, in their turn, reflect idiosyncratic facts about our lives. Supposing that $m_1 \ldots . m_n$ are the molecules composing the vase on your coffee table, if we were much smaller, so that vases were to us what, in actuality, galaxies are to us now, whether $m_1 \ldots . m_n$ constituted a single object would be a function of the explanatory context. Similarly, if we were much bigger and the rate at which we uploaded and processed information were much slower, we might see galaxies in much the same way that, in actuality, we see vases—we might see them (wrongly) as being as unconditionally thingy, since, relative to what we'd need to know under such circumstances, it would (given certain additional assumptions as to what our interests and needs would be, were we such slow-thinking giants) only be in hypothetical or, at least, rather exotic contexts that they lacked the properties requisite for thing-status.

6.3.4 The solution to P_2

These facts about the concept of thinghood help us solve P_2. It's meaningless to ask 'how far can m stray from *C*'s periphery---i.e., from the outermost layer of molecules that are *definitely* constitutive of *C*---without ceasing to belong to *C*?' The reason it's meaningless is that whether *m* is, or is not, a part of *C* depends on the explanatory context. Relative to certain questions about *C*, *m* would definitely fail to qualify as part of *C* were it so much as one tenth of an inch outside of the aggregate of molecules that

definitely qualified as being identical with, or very close to, *C*'s surface. (We'll henceforth refer to that molecule-aggregate as "*CS*"—short for "*C*'s surface.") In other contexts, *m* would definitely be a part of *C* if it were separated by a distance of 10 inches (or less) from *CS*. (Remember that, in this context, we're assuming, as to simplify the discussion, that *m*'s distance from *S* is the only thing that matters, so far as *m*'s being a constituent of *C* is concerned. In all likelihood, other things would matter; e.g., the manner in which *m* was moving, the frequency with which it collided with indisputably *C*-constitutive molecules and the manner in which such collisions occurred, etc. But analogues of the points we're making hold with respect to the question "how often must *m* collide with molecules that are clearly *C*-constitutive?" or "with how much force and in what exact manner must such collisions occur?" etc.)

In contexts of either of these two kinds, there would be some number *n* such that, if *m*'s distance from *S* were *n* or less, *m* would determinately belong to *C* and such that if *m*'s distance from *S* were *n* + 1, *m* would definitely fail to qualify as a member of *C*. In contexts of the third kind, this would not be the case. But that is because, in such contexts, "*C*" would have no definite referent. It is *not* because, owing to some "objective" indeterminacy---some gap or blank in the world---there was no fact as to whether or not *m* was a part of *C*. The real reason would be as follows.

6.3.5 An objection to our solution

"But aren't there contexts," it will asked, "where, relative to the question of whether *m* was or was not a constituent of *C*, it wouldn't matter by exactly how much m was separated from *CS*? And, supposing that such contexts exist---which they presumably do----doesn't that show that there is objective

indeterminacy?"

Each of these questions is to be answered with a "no." Sentence-tokens that, in terms of their surface-structure (what they sound like when spoken, how they look when written down) have the form "*x* is a constituent of *C*" are elliptical for statements of the form "relative to question *Q*, *x* is a part of *C*." And sentence-tokens of the latter kind are elliptical for: "the supposition that, when its distance from *S* is *n*, *x* is a part of *C* makes it easier to give a predictively/explanatorily useful answer to *Q* than does the negation of that supposition." Until a question about *C* is identified, the question "is *m* still a part of *C* when it's distance *n* from *S*?" is as meaningless as the question "is Topeka to the north of?"[495], and meaningless for the same reason. For argument's sake, suppose that, given some question *Q* concerning *C*, it doesn't matter, from the viewpoint of producing a predictively/explanatorily good answer to *Q*, exactly how far *m* is from *S*. In that case, the statement "*m* is a part of *C*" is *ipso facto* false, since "*m* is a part of *C*," if meaningful, is elliptical for "there is some salient question about *C* such that it is (relatively) easy/hard to answer that question properly by supposing that *m* is (not) a part of *C*."

It's true, of course, that if we *just* consider some cloud and some molecule on its periphery--- in other words, that if, without having any question in mind (other than whether that molecule belongs to that cloud), we ask "is that molecule a part of that cloud?"---we draw a blank and forthwith succumb to the temptation of thinking that, for some *x* (namely, that molecule), and for some property *P* (namely, the property of being a part of that cloud), that there is no fact as to whether the proposition *x is an instance of P* is true or not. But that's because the expression "the property of belonging to *C*" doesn't denote a property *except* relative to some question about *C*. A consequence is that, in the absence of such a question, the

occurrence of “the property of being a constituent of *C*” in “molecule *m* is an instance of the property of being a member of *C*” is functioning as a free variable. A consequence of this consequence is that, in the absence of such a question, “molecule *m* has the property of being a constituent of *C*”—or, more simply, “*m* is a part of *C*”—is a sentence-form, not a sentence proper, and therefore expresses no proposition at all.

One point should be made clear. To say that there is no fact as to whether a proposition is true or not is not to say that it’s definitely in some no-man’s-land between truth and falsehood. For any proposition in such a no-man’s land would be definitely *not* true and definitely *not* false; and there *would* therefore be a fact as to whether it was true and also as to whether it was false.

6.4 A third kind of Sorites paradox

Let us now move onto the Sorites paradox in the form in which it was first presented; i.e., let us consider the so-called “Paradox of the Heap.” Sorites paradoxes of this kind are to be solved along lines similar to those along which those of the second kind are to be solved:

P_3: A single grain of sand is not a heap of sand. If *n* grains of sand don’t jointly constitute a heap, then neither do *n* + 1 grains of sand. In other words, adding a grain of sand to a non-heap won’t yield a heap. Therefore, a billion grains of sand don’t constitute a heap.

As it stands, P_3 isn’t really a paradox. First of all, a billion grains of sand, even when taken jointly, *don’t* necessarily constitute a heap. If a million grains of sand are dispersed through space---if they are ‘scattered’---they

don’t constitute a heap. So for a given multiplicity of grains of sand to constitute a heap, any given grain must be (relatively[496]) contiguous with at least one other member of that multiplicity. But even though, for a multiplicity of grains of sand to form a heap, it is necessary that they fulfill that condition, it isn’t sufficient. For a billion grains of sand that are arranged in a straight line, each contiguous with its neighbors (or, if it’s the first or last grain, with its one neighbor) isn’t a heap. For a multiplicity of grains of sand to qualify as a heap, it is necessary and (ultimately) sufficient that they satisfy the following requirements:

(i) each member of that heap must be contiguous with at least one other such member;

(ii) taken jointly, those grains of sand must form a 3-D entity (I’m using the generic term ‘entity,’ instead of the word ‘structure’ for reasons to be made clear right away);

and

(iii) the entity in question cannot be a *structure*, at least not in the same sense as a building or a ship in a bottle.

(i) and (ii) are self-explanatory, but (iii) requires clarification. A sand-*castle* is not a heap of sand. What is the relevant difference? “Sand-castles are products of human artifice, whereas heaps of sand are not.”

First of all, heaps of sand are often products of human artifice; second of

all, there are structures (e.g., plants) that are not the products of artifice that are definitely heaps and wouldn't be heaps if, other things being equal, they were made of sand. The difference between a heap of sand and a sand-castle seems to lie in the fact that there are quite strict conditions that a grain of sand must satisfy to become part of a sand-castle, whereas there are no conditions, apart from some very generic ones, that a grain of sand must satisfy to become part of a heap of sand. To become part of a given sand-castle, a given grain of sand must be affixed to the sand-castle in such a way that it provides extra support for the walkway connecting the two towers, or that it completes some otherwise incomplete decoration on the castle, etc. But to become part of a heap of sand, a given grain of sand need only be adjacent with one or more of that heap's existing constituents.

These points make it clear how to solve a paradox that is similar to P_3 and that, as we'll see, is to be solved in much the same way:

(P_3#) A single grain of sand is not a heap. If *n* grains of sand, no matter how they are arranged, don't jointly constitute a heap, then adding another grain to those grains won't yield a heap, no matter how that extra grain is added to that heap.

What makes it seem that (P_3#) is hard to solve that the word "heap" hasn't been defined. I don't mean that it hasn't been *precisified*, but that it hasn't been disambiguated. If a "heap" refers to a *structure*, as opposed to an amorphous mound, of a certain kind, then, depending on what sort of structure that is, adding one more grain of sand to a structure consisting of n grains of sand quite definitely will turn a non-heap into a heap.

If by a "heap," we mean an amorphous mound of some kind, then, given

a few fairly innocuous points about language that we've already made, we've already solved P_3. Let us (speaking as logicians, not as semanticists) regard "heap" as a noun form of the adjective "heapy." There are obviously degrees of heapiness---just as there are degrees of intelligence, tallness, etc. Whether a given mound of sand is "heapy" is relative to the operative standard. Let A and B consist of n and $n + 1$ grains of sand, respectively, but otherwise be qualitatively identical. To an ant whose objective is make it over mound A, A may not be heapy---meaning that the ant can make it over. But B may be heapy to that same ant, given that, because B contains that extra grain of sand (as to whose position in B I am making definite assumptions, clearly), B is not surmountable. So relative to the ant's objectives (and also, perhaps, relative to the ways it can think of surmounting a heap), B is, whereas A is not, heapy.

6.4.1 Summing up

P_3 is dealt with in two steps.

(1) Replace "heap" with the corresponding adjective.

(2) See that adjective as having (hidden) relational structure.

Logical vs. statistical vs. explanatory probability

Supposing that there are 10 balls in the urn, and only one of them is white, it follows, relative to at least one disambiguation of the word "probability," that:

(WB) The probability of an arbitrary one of them being white is 1 in 10.

Given that all unsupported terrestrial objects (setting aside airplanes, helium balloons, and other objects subject to forces more powerful than, and acting in opposition to, that of gravity) have fallen so far, it follows that:

(UB) The probability that future unsupported objects (setting aside helium balloons, etc.) will fall is extremely high.

It's unclear whether makes sense to assign a numerical value to that probability. But, supposing that it does, it is presumably in excess of 99%.

Given the data (e.g. the Michelson-Morley experiment, the anomalies in Mercury's orbit),

(RB) The probability that Special Relativity Theory is correct is exceedingly high.

Echoing what we said a moment ago, given the doubtful supposition that it makes sense to assign a numerical value to the just-mentioned probability,

it is presumably in excess of 99%.

On the face of it, it seems reasonable to say that the occurrences in WB, UB, and RB of the word "probability" refer, respectively, to *logical*, *statistical*, *explanatory* probability. And *prima facie* it also seems that no one of these three forms of probability can be absorbed into, or identified with, either of the other two forms of it. In other words, logical probability is logical probability is *sui generis* and thus incapable of being identified with, or understood in terms of, either explanatory or statistical probability, the same thing *mutatis mutandis* being true of each of the latter two kinds of probability.

According to the statistical interpretation of probability, to say that it's likely that Smith will have 9 more drinks tonight, given that he's just had his first drink this evening, is to say that, if C is the class of evenings on which Smith has at least one drink, more than half of C's members are evenings on which Smith has had ten drinks. Setting aside a few empiricist hard-liners (e.g. Mill, Carnap), whose views on this matter are now universally rejected, philosophers have either denied, or refrained from affirming, that the sort of probability mentioned in RB can be assimilated into, or can itself assimilate, the sort mentioned in UB.

I agree with the conventional wisdom that explanatory probability cannot be absorbed into statistical probability or, of course, logical probability. But I *do* believe that statistical probability can and must be absorbed into explanatory probability---that the concept of statistical probability, if it isn't completely incoherent, is but one component or facet of the concept of explanatory probability. So-called "statistical probability," I will argue, is really explanatory probability in disguise.

I will further argue that statistical probability and logical probability aren't identical or even similar, the second word in WB being nothing more

than a homonym of the second word in UB.

Underlying my defense of these claims is a third, which I will defend right away, namely, that enumerative induction, supposing it to have any merit as a source of knowledge, is a form of (or, more accurately, is a component of) inference to the best explanation

Induction by enumeration vs. inference to the best explanation

Given only that every single one of the 5,000 first editions you've ever come across smelled like pipe tobacco, you cannot reasonably infer that all first editions smell like pipe tobacco. What if all of those first editions were the property of your pipe-smoking friend Larry? In that case, you'd have good to reason to suspect that their smelling like pipe tobacco had to do, not with their being first editions, but with their being things that had been around Larry. By the same token, so far as you are entitled to infer that all first editions smell like pipe tobacco, it's only to the extent that you have reason to believe that a thing's smelling like pipe tobacco is *rooted* in its being a first edition. But that means that, if the data entitles you to infer that all first editions smell like pipe tobacco, it's only because it also entitles you to accept some inference to the best explanation (one to the effect that a thing's being a first edition is, for some reason or other, responsible for its smelling like pipe tobacco). In general, induction by enumeration is parasitic on inference to the best explanation.

Another example: You are playing a game of roulette. You know that nothing about the game is in any way rigged and that that there is nothing about you--- your body, mind, or dice-throwing practices---that disposes a die that you've thrown to roll one way as opposed to anotherYou've rolled the dice seven times. Each time, at least one of the die has landed in slot #27. Under these circumstances, it would be natural to predict that you'll roll a 27 next time. But that prediction would be irrational: you have no more reason to believe that you'll roll a 27 than you will any other number. By the same token, no matter how many consecutive 27s you roll, you are never any more entitled to bet that, next time, you *won't* roll a 27 than you are to bet that you won't roll any one of the other available numbers. Since, by hypothesis, the

conditions that obtain on any two occasions are identical, it would be irrational on any given occasion to think any one outcome any more or any less likely than any other.

When *would* it become rational to believe that, next time, you're more likely than not to roll this as opposed to that number—that, for example, you're especially likely to roll a 27? This belief becomes rational when, and *only* when, you have reason to believe that a 27-roll is favored by the structures involved in the game. It would be rational were to you learn that the dice are magnetically attracted to the 27-slot, or that the 27-slot is much bigger than any of the other slots. So a long-run of 27's *is* evidence of another 27 *to the extent* that it's evidence of a mechanism that programs for a 27. In and of itself, a long of 27's is, as Hume made clear, evidence of nothing. Supposing that (i) no roll affects any later and (ii) that there's nothing underlying the various rolls that directs them towards a common outcome, the fact that you've rolled consecutive 27's is *ipso facto* irrelevant to what you'll roll next.

The best, and the only, explanation of the fact that all known phi's are psi's is that, thanks to some mechanism, a thing's being a phi disposes it to be a psi; and as soon as such a mechanism has been posited, and not one second sooner, your knowing of a given thing that it's a phi is a reason for you to hold that it's a psi. So any non-fallacious "induction by enumeration" is parasitic on a case of inference to the best explanation.

A consequence According to the frequency interpretation of probability, for it to be probable that x has psi, given that x has phi, is for phi's that are psi's to outnumber phi's that are non-psi's. Given what we've just said, this must be qualified. Supposing the number of phi's that are psi's to be high (relative to the number of phi's that are non-psi's), that is evidence that the next phi

will be a psi *only* to the extent that it's evidence of something that disposes phi's to be psi's.

A small but heterogeneous class of phi's that are psi's may definitively establish that that all phi's are psi's and a large but homogenous class of phi's that are psi's may provide no support for it.[497] Why does a heterogeneous reference class matter? If two otherwise dissimilar phi's are psi's, we know that, if anything about either is responsible for its having psi, it is its having phi. But if two otherwise similar phi's are psi's, we don't (for that reason alone) have any reason to conclude that either one's being a psi has anything to do with its being a phi.

Thus, the more resilient the connection is between a thing's being a phi and its being a psi---the more varied are the contexts within which a thing's being a phi is accompanied by its being a psi---the better the chance that the next known phi will be a psi. Presumably this is because causal relations do, whereas mere correlations do not, tend to be context-independent.

Also, there are many perfect concomitances that, for reasons having nothing to do with their degree of context-invariance, we instinctively know to have no predictive or explanatory importance. We know that every green emerald is grue (green if examined before t (today's date) or blue and examined thereafter); but we don't think that tomorrow emeralds will be blue. Be it noted that the property of being an emerald no less resiliently, context-invariantly associated with the property of being grue than with the property of being green. So there is no reason to suppose that the next phi will be *given* only the fact that, in each of many qualitatively different contexts, they are frequently (always) conjoined. Given that fact *with* some other fact, which we have to identify, it *is* likely that the next phi will be a psi. But this suggests that concomitances have no predictive value, save what they leach

off of something underlying them.

Why is so little predictive stock to be put in the fact that all known emeralds are grue? xxx

This is made clear by the following paradox, commonly referred to as the "lottery paradox" ("LP," for short):

(LP) There's a lottery. One million tickets are sold. There's one winner. So, given any one ticket, the probability of its losing are 999,999 in 1,000,000. You buy a ticket. Let *x* be this ticket. You know that ("LT" is short for "losing ticket"):

(LT) the probability of ticket *x*'s losing is 999,999 in 1,000,000.

But do you *know* that it'll lose? No. It may even have been rational for you to buy the ticket. Given what you stand to lose by buying the ticket and given what you stand to gain in the remote chance that you win, it may have been in your interest to buy it, even after you take into account that the chances of your losing is 999,999 in 1,000,000. But this decision couldn't possibly have been rational if you *knew* that your ticket would lose.

In general, there's no knowledge where there isn't certainty. If, given the information at your disposal, there is *any* chance, no matter how small, that *P* isn't the case, then you don't *know P*. We all know (or so let us suppose) that:

(UB) the probability that future unsupported objects will fall is extremely high, given that such objects have fallen so far.

Is there *any* chance that, all of a sudden, unsupported objects will no longer fall? Yes. (The statement "all of a sudden, unsupported objects will no longer fall" isn't ruled out by the laws of logic. So it could, theoretically, be true.) Therefore, you don't *know* that they won't fall; and, by obvious extensions of this arguments, you don't *know* anything about the future (apart from sterile trivialities; e.g., "whatever happens, happens"). The very concept of "probable" knowledge is incoherent. The merely probable is *ipso facto* not known. To know it to be 99.9999% likely that unsupported objects will continue to fall is to *not* know that they'll fall.

The problem with LP

LP uses word "probability" in an equivocal manner. Once this is seen, LP loses any appearance of cogency that it might otherwise have had.

Let's take another look at:

(LT) the probability of ticket *x*'s losing is 999,999 in 1,000,000.

LT says that there are one million tickets but only one winner. That's all it says. But UB doesn't say anything of the sort. UB doesn't make a *statistical* point. It doesn't say: "nine times out of ten, unsupported objects fall." UB makes a point about the relative merits of two competing hypotheses, these being:

(H_1) unsupported objects will continue to fall,

And:

(H_2) unsupported objects will not continue to fall.

UB is true if, given the data that we have, H_1' s being true would undermine the integrity of our belief system to a lesser degree than H_2' s being true. Given the information that we have right now:

(SB)[498] there is only an extremely low probability that walking on stilts for an hour while wearing a beanie with a propeller on it is a foolproof cure for cancer.

Does SB say that better than nine cases of cancer out of 10 *aren't* cured in this way? No. It doesn't make *any* sort of statistical claim. SB is a claim, not about statistics, but about explanatory integrity. It says that more damage will be done to our belief system if it turns out that cancer *can* be cured in that way than if it turns out that it cannot. If, by some remote chance, cancer *can* be cured in that way, everything we "know" about medicine and pathology is false. But the damage wouldn't end there. Our medical beliefs are rooted in general assumptions about the causal structure of the world. (For example, given the information that Smith has gone into anaphylactic shock whenever he's eaten peanuts or anything with peanuts in it, even cutting-edge medical researchers will assume that Smith will go into anaphylactic shock if he eats peanuts that have been painted green. This assumption is rooted in generally held causal beliefs. No inside information is involved.) Were it found out that SB was false, many of the commonsense-based underpinnings of our medical beliefs would have to be reconsidered. But all of those beliefs remain intact if cancer cannot be cured in that way. It is improbable that walking on stilts (etc.) cures cancer to the extent that our current beliefs depend on its *not* being a cure. And it's probable that walking

on stilts (etc.) isn't a cancer cure to the extent our beliefs depend on it.

In general, *P* is probable to the extent that our beliefs depend on it, and improbable to the extent that they depend on its negation.

Counterstatistical vs. counterexplanatory

Counter*statistical* events are innocuous, but counter-*explanatory* events are not. Smith rolls double-sixes five times in a row. Statistically unlikely—yes. Important—no. Must we change our belief system because of it? No. Need we reconsider our assumptions about the basic structure of the world? No. And it makes no difference whether it's five double-sixes in a row or a billion. The improbable sometimes happens, and that's the end of it. When it happens, we have a nice anecdote for cocktail parties; we don't have a situation worthy of scientific scrutiny. But we *do* have such a situation if it turns out that Smith is using telekinetic powers to make the dice come up sixes. For, in *that* case, some deeply held convictions of ours have to be revised.

In describing Smith's rolling double-sixes five times in a row as "improbable," what one is saying is very different from what one would be saying in describing Smith's having telekinetic powers as "improbable." In the first case, one is affirming a mere statistical improbability. One is saying only that such things don't *often* happen. In the second case, one isn't saying anything of the sort. One is saying Smith's having such powers is inconsistent with some fundamental beliefs of ours—with beliefs that, if jettisoned, would take a lot of other obviously (or at least presumably) correct beliefs with them.

Given *any* event, there is a way of describing it that makes it improbable in the *statistical* sense. In fact, *any* given event, described at all precisely, has

a statistical probability of vanishingly close to zero. I flip a coin. Nothing improbable about this. But it's a safe bet that no coin has ever moved in *exactly* the way this one did after I flipped it. Described as a case of my flipping a coin, the event isn't improbable at all.

But that same event is extremely improbable if described as a case of an object's moving in manner M, where M is the exact way that it did move. But even though nothing apart from that coin has ever moved in manner M, that coin's doing so isn't a matter that deserves the slightest scientific scrutiny. Its moving in that way is consistent with everything we know about the world, and there is therefore nothing to be explained. But *there is* something to be explained if Smith turns out to have telekinetic powers, since his having such powers is improbable in the *explanatory* sense. It's irrelevant that it is also improbable in the statistical sense. *Every* event is improbable in the statistical sense.

My winning the lottery is improbable in the statistical sense, but not in the *explanatory* sense. And that's why, before the winner is announced, we can't say that we *know* it won't win: my winning is consistent with everything we know about the world. My sprouting wings in the next five minutes is improbable in the *explanatory* sense. And that's why we *can* say that we *know* it won't happen. It's *in*consistent with everything (or much, at any rate) that we know about the world.

Grue: the problem of counterinduction

A puzzle devised by Nelson Goodman[499] will clarify, as well as support, what we've said:

(GA)[500]: Let's invent a new word: "grue," to be defined as follows. Something is grue if it is green before Jan. 1, 2010, or blue thereafter. So a shirt that's green on Oct. 12, 2007, is grue, as is a shirt that's blue on Nov. 11, 2012. Let's assume that, before Jan. 1. 2010, all emeralds have been green. (From now on, we'll abbreviate "Jan. 1, 2010," as "2010.") In that case, all emeralds have been grue prior to 2010. Anything that is green-and-examined-before-2010 is also grue. So the fact that all emeralds examined before 2010 are green means that all emeralds examined before 2010 are grue. We are assuming for argument's sake that, given the fact that all emeralds examined before 2010 were green, we are inductively warranted in believing that all emeralds examined after 2010 are green. By that logic, it follows that, given the fact that all emeralds examined before 2010 were grue, we are inductively warranted in believing that all emeralds examined after 2010 are grue. But anything examined after 2010 that is grue is *blue,* not *green.* So, given that all emeralds before 2010 were green, our inductive grounds for believing they'll be *blue* after 2010 are as strong as our grounds that they'll be green.

The degree to which a given fact is *statistically* improbable depends on how it is described. Given some emerald *x* that was examined before 2010 and found to be green, "*x* is grue after 2010" is, from a strictly *statistical* viewpoint, equiprobable with "*x* is green after 2010" and also with "*x* is blue after 2010." But that's no cause for alarm. Statistical improbability is innocuous. It's explanatory improbability that you have to worry about. But the explanatory probability of "*x* is green after 2010" dwarfs that of "*x* is grue after 2010" and, therefore, "*x* is blue after 2010." The supposition that objects don't change color (except under conditions that emeralds don't satisfy, at least not in virtue of making it to 2010) is deeply embedded in scientific

theory and also in the commonsense underpinning of scientific theory. If it's wrong, then science is in serious trouble, and so is human knowledge in general.

In response to this, it will be said that:

(OA)[501] If GA is right, emeralds that "turn blue" at the beginning of 2010 *don't* change color. They stay the *same color*—grue. And emeralds that "stay green" *do* change color, since they go from being grue to having some other color. (Or, to be more precise, emeralds *do* change color at the beginning of 2010—but only *relative* to one set of color-terms ["red," "green," etc.]. But emeralds stay the same color *relative* to different color terms ["grue," "bleen," etc.].)

OA simply isn't true. Pictures of green emeralds (taken before 2010) show them *not* to have the same color as blue things after 2010. Grueness and greenness are different *properties.* But they're not different *colors*. Grueness isn't a color at all—it's a color-*related* property. Let's say that *x* is a "squiangle" (rhymes with "triangle") if it's square-shaped before 2010 and triangle-shaped thereafter. If, for some reason, some object that was square before 2010 suddenly becomes a triangle, it's obviously changed shape. It used to have four sides. Now it has three. And a "squiangular" object that is square-shaped *after* 2010 has started—in other words, an object that had four straight sides, etc., before 2010, and continued to have that shape—hasn't *changed* shape; it hasn't gone from being a "squiangle" to being a square. It's stayed the same shape. The same goes for "grue."

[1] Advocates of this argument say that the true sentences refer to *The True*, and that false sentences refer to *The False*. They don't say what these terms mean. In an effort to free their argument of needless obscurities and indefensible assumptions, I've taken the liberty of assuming that they refer to the properties of being true and false, respectively.

[2] Give or take some irrelevant details, this argument was put forth by Frege in his landmark paper *On sense and reference*.

[3] Inevitably, some will quibble over whether, when they are interpreted in this (Russellian) fashion, definite descriptions are quantifiers. First of all, it doesn't matter. If, when read attributively, the meaning of "the inventor of bifocals is smart" is identical, or equivalent, with the proposition something or other uniquely invented bifocals and any such thing is smart, then, thus read, that sentence is an existence-claim; in which case, the occurrence of the "inventor of bifocals" therein isn't a singular term and, at the very least, quantifier-*like*. Second, when read attributively, definite descriptions *are* functioning as quantifiers.

But, for what it's worth, definite descriptions *are* quantifiers *if* Russell is right about them. A quantifier is an expression that, when coupled with a predicate yields a sentence saying how many things of a given kind have that property. "All people" is a quantifier, since, when coupled with "snore," which denotes the property of snoring, yields "all people snore," which says how many things of a certain kind have the property of snoring. "No fish" is a quantifier since, given the expression snore, yields the sentence "no fish read Shakespeare," which says how many things of a given kind have the property of reading Shakespeare.

Thus "no fish" is to be defined thus: in virtue of having the form "…no fish…," a sentence says that the class of fish doesn't overlap with the class of things x such that…x…("No fish snore" says that the class of fish doesn't overlap with the class of things x such that x snores. "Jim hates no fish" says that the class of fish doesn't overlap with the class of things x such that x snores.) "All people" is defined thus: in virtue of having the form "…all people…," a sentence says that the class of people is a subset of the class of things x such that…x…("All people snore" says that the class of people is a subset of the class of things x such that x snores. "Jim hates all people" says that that the class of people is a subset of the class of things x such that Jim hates x.)

Assuming for argument's sake that the CBR is the proposition meant by CB, the semantic rule for "the inventor of bifocals" is given by the proposition that, in virtue of having the form "…the inventor of bifocals….", a sentence says that the class of things that uniquely invented bifocals has exactly one member in common with the class of things x such that…x…A consequence is that, given the aforementioned assumption, "the inventor of bifocals" is a quantifier. Given any other

definite description, the same argument *mutatis mutandis* shows that, when read attributively, it is functioning as a quantifier.

[4] What follows is a duly updated version of Footnote 3. Frege said that true sentence refer to *the True*. But he didn't say what *the True* is, and it's a matter of debate what exactly *the True* is supposed to be. But whatever *the True* is, let's stipulate, strictly for argument's sake, that "the property of being true" refers to that thing. If, in the forthcoming argument, each occurrence of "the property of being true" replaced with E, the resulting argument is cogent iff the original argument is cogent.

[5] A less vivid, but more accurate (and also more general), formulation would be: *R refers to O exactly if, in virtue of having the form* ⌜*...R...*⌝*, a sentence S is to the effect that...O...*

[6] To clarify the parenthetical qualification: given what is meant by “bald”, “is” and all of the material occurring in that sentence besides “Benjamin Franklin”, if you couldn’t attribute baldness to BF by saying “Benjamin Franklin is bald”, then, in that context at least, “Benjamin Franklin” wouldn’t refer to BF. Why include this qualification? Because there could obviously exist a language in which “bald” referred to BF and in which some homonym of the “Benjamin Franklin is bald” attributed baldness to BF. So given *only* the supposition that one can’t attribute baldness to BF by saying “Benjamin Franklin is bald”, it doesn’t follow that “Benjamin Franklin” doesn’t refer to BF. But this conclusion does if follow *given* what the remaining semantic (and, presumably, syntactic) rules of English are.

[7] Strictly speaking, there is no such thing as *the* sentence "snow is white." There is the sentence-type; and there are tokens of that sentence. And, for reasons that I've identified elsewhere (*note to editor: I've omitted the references since the present paper is still being reviewed—I'll supply them upon your request*) that sentence-type is not semantically comparable, at least not in any respects that are relevant here, to its tokens.

Also, it seems to me that, in this context, it's important to distinguish between embedded occurrences of "snow is white" (an utterance of that sentence that is constitutive of an utterance of some other sentence, e.g. the occurrence of it in a given utterance of "if snow is white, then at least one substance is white") and non-embedded occurrences of that sentence. There is a difference between saying "snow is white", on the one hand, and "the proposition that snow is white." A non-embedded utterance of the latter is neither true nor false. A non-embedded utterance of the former *is* true or false (since it's true). And if, having been ordered to identify a proposition that attributes a color to some substance, I say "the proposition that snow is white", I've clearly done what I've been told. But if, having been given that same order, I say "snow is white", I haven't *clearly* done what I've been told. This suggests that, if it isn't outright false, the thesis that "snow is white" refers to the corresponding proposition must be heavily qualified.

And even when it occurs as a proper part of another sentence, "snow is white" isn't interchangeable with "the proposition that snow is white." Whereas " grass is green and snow is white" is true, "grass is green and the proposition that snow is white" is ill-formed and is therefore neither true nor false.

That said, the semantic gulf between "grass is green and snow is white" and "grass is green and the proposition that snow is white" is miniscule compared to the semantic gulf between "grass is green and snow is white" and "grass is green and the property of being true." And

My feeling is thus that both embedded and non-embedded occurrences of "snow is white" *do* refer to the proposition that snow is white---but that this isn't *all* that they do. I suspect that such

occurrences contained phonetically unrealized material *not* contained by "the proposition that snow is white"; and it is because of these (alleged) differences, I suspect, that "snow is white" isn't, at least not categorically, interchangeable with "the proposition that snow is white."

[8] See his papers "Truth and meaning" and "True to the facts," both of which were published in his (1984) collection of his papers *Essays on Truth and Interpretation.*

[9] Given a vague expression E, a 'precisification' of E is an assignment of meaning to E that, although non-vague, is otherwise like E's existing meaning. See Section 1.2. The term 'bald' is vague. But we'd be precisifying the word "bald" were to say exactly how many hairs (of a certain thickness) a person x (whose cranium was of a certain size) had to have for ⌜x is bald⌝, x is bald to be true. (It seems to me that this sort of precisification may not be in quite the same category as the precisifications put forth by science (e.g. 'temperature=mean kinetic energy'). There are (at least) two reasons for this. First, scientific precisifications are less stipulative, and more descriptive, in nature than precisifications of the sort just discussed. Second, whereas non-trivial, non-linguistic discoveries must be made to successfully precisify terms such as "temperature" and "acid", such discoveries are irrelevant when it comes to precisifying terms like "bald" and "nice": the requisite knowledge is of a highly generic and, for the most part, strictly linguistic nature.

[10] I'm including *mental* (mass-)energy. Of course, materialists will say that, since mind is matter, it's unnecessary of me to say this. But I don't want to prejudge the question whether the mental is identical (or otherwise constituted by) the physical. (Also, speaking personally, even though I am, ultimately, a materialist, I have dualistic sympathies.)

[11] Oxford University Press, 2005.

[12] In what follows, I use actual quotation marks where I should, strictly speaking, use quasi-quotation marks.

[13] In what follows, we'll set aside irrelevant subtleties relating to the tense-markers occurring in the sentences we're examining, and we'll also set aside subtleties relating to the fact that BF is no longer with us.

[14] Why do I say "seems"? CB is a sentence. Given only that CB is equivalent with CBR, it doesn't follow that the proposition meant by the one is equivalent with the proposition meant by the other. Strawson (1950) made this clear. (Strawson's insight was popularized by Kaplan (1989.) Suppose for argument's sake that the semantic rule for "that person over there is" is: if, in a context C, somebody O is uniquely a salient person, then an utterance in C of ⌜that person over there is has phi⌝ has for its meaning some proposition equivalent (if not identical) with *O has phi*; and if, in C, there is no uniquely salient person, then such an utterance is neither true nor false (the reason being that the demonstrative doesn't have a referent). But, given that same supposition, the *sentence* ⌜that person over there is has phi⌝ would be true exactly if, in the relevant context, somebody x were uniquely a contextually salient person and, moreover, any such person had phi.

Thus, given only that Russell's theory correctly models the logic of CB and other *sentences* containing definite descriptions, it doesn't follow that it correctly identifies the *propositions* meant by such sentences. My own view is that, indeed, TD is false. But I also believe, partly for the reasons given here, that Frege's view is false. My alternative to their analysis is put forth else where (*note to editor: I'll supply the references upon request).*

[15] "FR" is short for" "Frege."

[16] Frege (1892).

[17] " "IB" is short for "the inventor of bifocals."

[18] Actually, according to TD, each of (1) is and (2) is ambiguous between "wide-scope" and "narrow scope" interpretations, (1*) and (2*) being the narrow scope interpretations. (See Russell 1905.) In any case, what's important is that (1*) and (2*) are among the meanings that TD ascribes to (1) and (2), respectively.

[19] Special compositionality is to be distinguished from the more general principle that the meaning (semantics) of a given expression is a function of the meanings (semantics) of its constituents.

[20] In any case, supposing that O is the prettiest woman in existence, it refers, according to Frege, to *something* that is individuative of O. It doesn't matter, at least not in this context, whether, when referring to that thing, we use the term "concept" or "property."

[21] In his paper "Theories of meaning and learnable languages", Davidson masterfully shows why, if Frege's semantics is correct, no human language would be learnable. That paper is published in *Inquiries into Meaning and Truth*. Also published in that volume is Davidson's paper "True to the facts", in which he argues that there is only one fact. This argument of his is discussed in Section 1.3 of the present paper. That argument, it will be recalled, is a variant of SS. And, as Frege himself realized, a corollary of Frege's semantics is that SS is cogent. Thus, Davidson himself showed that Frege's semantics is fallacious; and he does this in the very same book where, on the basis of SS, he tries to show that there is, at most, one fact. SS assumes the truth of Frege's semantics and, therefore, of the very semantics on which Davidson's one-fact (or no-fact) argument is based. In that same book, Davidson puts forth an entire philosophical system that is, in part, based on the assumption that there is, at most, only one fact. It follows that, in one and the same book, puts forth a philosophical system and refutes that same system.

[22] See e.g. Dummett (1978) Fine (1999) , Williamson (1994, 1998a, 1998b), Williamson and Sainsbury (1998), Sainsbury (1999), Wright (1999), Parsons (2000), Sorensen (1998), Cargile (1999), and Tye (1999). The view put forth in the present paper is not even mentioned, let alone considered, in any of these works, or for that matter in any of the works known to me on vagueness or the Sorites paradox.

For a review of all the dominant solutions to the Sorites, see Williamson and Sainsbury (1998). The solution I will present is not among those considered by Williamson and Sainsbury.

[23] See Williamson and Sainsbury (1998: 458). I am relying on Williamson and Sainsbury for this historical information.

[24] See Williamson and Sainsbury (1998: 466-481). I say that *almost* all of the solutions involve an abandonment of some law of classical logic; for there is the so-called epistemicist view, according to which *there is* a sharp cut-off between being bald and not-bald, rich and not rich. So somebody with $10,000 is not rich, while somebody with $10,001 is rich. But such a view seems deeply arbitrary; and 'there is no evidence that [predicates like 'rich' and 'bald'] induce sharp cut-offs.' (Williamson and Sainsbury, p. 479.) Also, it is hard to believe that rich has such a *precise* meaning, given that absolutely no one *knows* what that meaning is: but surely since it is *people* who give meaning to terms like 'rich' and 'bald' through their use of such terms, it makes no sense to suppose that people should be so radically ignorant of the conditions of their application. (Williamson and Sainsbury (1998: 480) make essentially this point.) So the epistemicist view is, understandably, not taken seriously.

[25] It is true that we are unlikely to characterize someone who is only marginally more intelligent than average as 'smart compared to the average person.' But that has to do with pragmatics, not with semantics. The margin of difference would be negligible and would thus be of no pragmatic significance (except in rare contexts, like the present one, where delicate points about meaning are just what is at issue). A comparison might help. For something to be 'massive' is, technically, just for it to have mass. But no one would ever describe a mouse as 'massive' (except in comparison to, say, a proton); for it would be very misleading to do so, even though such a characterization would technically be correct. Similarly, it would be misleading – though technically correct – to say that George was smart 'compared to average person' so long as George's IQ is so much as one point higher than average. In this paper, we are concerned with the finer aspects of meaning, not with the course-grained analogues of them constitute the bulk of everyday discourse. So we will use the expression 'compared to' in its strict and literal meaning; we will say that George is 'smart compared to the average person' iff George is smart*er* than the average person, no matter by how little. I further discuss this point at the end of the present paper.

[26] In an earlier draft of this paper, I confused sentence-negation with predicate-negation. I'd like to thank an anonymous reviewer at this journal for setting me straight on this.

[27] Donald Davidson makes this point, and uses it to good advantage, in 'The Logical Form of Action Sentences.'

[28] What follows is a response to a point made by an anonymous reviewer at *History and Philosophy of Logic.*

[29] Though I doubt it. 'x is smart compared to y' is quite obviously, in my view, merely a pedantic way of saying 'x is smarter than y.'

[30] Actually 'smarter' is *itself* such a predicate. Surely it is not a phonemic accident that the sound 'smart' occurs at the beginning of 'smarter.'

[31] Of course, Carl. Hempel (1945a, 1945b) discovered the paradox. The paradox has been re-articulated many times, sometimes with interesting twists. (See Maher 1999). Hempel bit the bullet, saying that one *does* confirm (R1) by encountering non-black non-ravens. An intuitively more plausible view is proposed by Quine (1969). (My solution here has some vague affinities with Quine's view.) Some sharp -- in fact, devastating -- criticisms of the views prevailing during his time, including Quine's, is given by Ayer (1971). Scheffler (1981) admirably reviews all the major solutions prevailing at the time of his writing. Most solutions propose that (R1) and (R2) are, indeed, equivalent. See Schlesinger (1974), Trusted (1979), Cohen (1989), Leavitt (1996), Sainsbury (1997), Maher (1999).

There are views that try to destroy the alleged equivalence between (R1) and (R2). But these views always resort to non-classical logics. (See Sylvan and Nola (1991)).So in effect these views *do* concede that, within a classical framework, (R1) and (R2) are indeed, equivalent. So these views presuppose that (R1) and (R2) *are* equivalent, if only within a classical framework: otherwise non-standard logics wouldn't be invoked to block the equivalence.

These are all worthy views. For reasons of space I cannot review them all here. But what I can say is that, since they all presuppose a certain equivalence between (R1) and (R2), they are hobbled at the outset. Again, I grant that on *a* legitimate reading, (R1) and (R2) are, within the framework of classical logic, logically equivalent. But "all" statements are not always so transparent in terms of their content. And this is a case where the nuances of the word "all" are relevant. I submit that once we make these nuances explicit, we will find that *even within a classical framework*, what is said (or at least meant or insinuated) by (R1) and (R2) is not equivalent.

[32] Hempel (1945a, 1945b), Sheffler (1981).

[33] Sylvan and Nola (1991)

[34] When I use the term “cause”, I mean to include partial causes. Obviously *total* causes are very hard to identify, and definitely are not to be found in the paradigms we are considering. When I say “x causes y”, I mean “x is causative of y”, not “x is causally sufficient for y” or even “x is causally necessary for y”. If you eat rotten fish, that is causative of your becoming ill. But it is not sufficient; if you had different digestive processes you wouldn’t have become ill. And it isn’t even necessary: you could have eaten rotten pork. But, in *some* legitimate sense of the word “cause”, your eating the fish is still a cause of your becoming ill. That is the sense in which “cause” is used in this paper. (It probably corresponds to Mackie’s (1977) INUS analysis. But we can leave that aside.)

[35] Notice that I say “nomic” and not (necessarily) “causal”. I distinguish between nomic and causal connections because – this is the received wisdom, anyway –some laws are not causal, but are merely “functional”. (See Mackie 1977.) Also, it seems quite right to say that there is a nomic relation between x’s having such and such temperature and x’s having thus and such mean kinetic energy. But the relationship is not causal; it is a relationship of realization or supervenience or, possibly, identity.

[36] See Goodman 1954, Hempel 1966.

[37] Again, not all the solutions take for granted the equivalence of R1 and R2. But those that do not invoke deviant or non-standard logics to block the equivalence. So such views *do* concede that, within classical logic, what is being said by (R1) is equivalent to what is being said by (R2). See Sylvan and Nola (1991).

The essence of my discussion is that, read the appropriate way, (R1) and (R2) are *not* equivalent: if we give the “all” the appropriate valence, there is *no* equivalence – even within the strictures of classical logic. So no non-classical logics are resorted to.

And this, surely, is a good thing. All other things being equal, it is usually best to try to solve a paradox within the framework of classical logic. Otherwise we run the risk of a global conceptual upheaval in our effort to deal with a relatively local one. We don’t want to induce systemic alterations of our conceptual system unless no alternative is possible. The solution here doesn’t involve systemic changes – it doesn’t involve any non-conservative logical moves.

[38] In the sense of being communicated, not necessarily in the sense of being literally or semantically meant.

[39] The word “consequence” here is *not* to be taken in the *material* sense. “snow is white” is a *material* consequence of “grass is green.” But the former is not a consequence, in any more robust sense, of the latter. In P4, the word “consequence” is to be taken as having a *robust* meaning. Granted, the word “consequence” has many different robust meanings: causal, metaphysical, logical. But it seems fair to say that, for one of those meanings, P4 is correct statement.

[40] The word “consequence” here is *not* to be taken in the *material* sense. “snow is white” is a *material* consequence of “grass is green.” But the former is not a consequence, in any more robust sense, of the latter. In P4, the word “consequence” is to be taken as having a *robust* meaning.

Granted, the word “consequence” has many different robust meanings: causal, metaphysical, logical. But it seems fair to say that, for one of those meanings, P4 is correct statement.

[41] I say “arguably” because, in my view, there is nothing obvious about the status of statements like “water is H2O”. We must distinguish the sentence from the proposition; and we must also distinguish the proposition from the information through which we access that proposition. Once these variables are taken into account, the obviousness that characterized discourse about the “a posteriori necessary” vanishes. I discuss this issue at length in my books__ and___.

[42] Of course, one could certainly *use* the sentence “all ravens are black” to say that, as it happens, everything that is a raven is also black. And there would be nothing wrong with such usage. But, it seems, when a zoologist – or, for that matter, practically anyone – uses that sentences, he/she means something a bit stronger than that every raven happens to be black. There seems to be some nomic content.

[43] The first to argue this was probably Berkeley, in *Principles of Human Knowledge.*

[44] The sub-script stands form ‘formal.’

[45] It would be more correct to say that (v) is *a* Russellian paraphrase of (iv). For ‘*There some x, such that given any y, y=x iff y is a king of France, and x punched Bob’* is also a Russellian paraphrase of (iv).

[46] Strawson is probably the most famous exponent of this view. See Strawson [1].

[47] See Blackburn 1984: 310.

[48] See Neale 1990: chapters 2-3.

[49] See Dummett 1981: chapter 1.

[50] Or if he were to utter a sentence of mathematics, logic etc.: tone-free sentences.

[51] Everybody who agrees that there are *any* completely synonymous sentences agrees that these are examples of that. Some philosophers, possibly Quine (at certain junctures) deny, in effect, that there is such a relation as synonymy, and *a fortiori* that such a relation is actually instantiated.

[52] MacCawley 1993: 302. See also Grice: 1989a.

[53] This exact point was made to me in private correspondence by Nathan Salmon.

[54] See Strawson 1990: 232-233. See also Graff: 2001 I borrowed this exampled from Graff.

[55][55] This that definite descriptions can be used predicatively, and the relevance of this fact to my paper, were brought to my attention by an anonymous reviewer at this journal. Addressing this point has considerably enriched this paper.

[56] Kripke 1991: 179.

[57] I have verified this with several linguists, including Tim Stowell (University of California, Santa Barbara) and Susanna Cummings (University of California, Santa Barbara).

[58] The contents of the next ten paragraphs are from Kaplan 1989.

[59] This is all neatly summarized by Kaplan in Kaplan 1989: 523.

[60] Again, I have verified this point with Professors Stowell and Cummings.

[61] The analysis I'm about to give is nearly enough given in Strawson's article.

[62] This problem was brought to my attention by an anonymous reviewer at this journal.

[63] The need to address this problem was brought to my attention by an anonymous reviewer at this journal.

[64] Kaplan 1989.

[65] Stroud (2002: 102-103).

[66] Examples of definite descriptions are 'the king of France', 'the table', and 'my neighbor'. A definite description might be defined as one that is *either* (i) of the form 'the phi' (where this expression is in the singular, not the plural, case) *or* (ii) as one that is synonymous with such an expression. So 'the king of France' is a definite description by (i). And 'my neighbor' is a definite description by (ii); for 'my neighbor' is surely synonymous with 'the person who is my neighbor.' The expression 'the people over there' is not a definite description, since it is in the plural case.

[67] The subscript stands for 'formal'.

[68] In this paper, by "proposition" I mean simply "sentence meaning." I am not using the term to denote some specific *conception* of what a proposition. Everybody agrees that a proposition is the meaning of a sentence. Where people disagree is as to what sort of a thing a sentence-meaning is. Russell thought that such things are complexes of objects and properties. Others think that they are sets of possible worlds or functions from worlds to truth-values. My use of the word "proposition" is neutral with respect to these different *theories* as to what a proposition is; I am, to repeat, using the term to mean "sentence meaning", not to denote this or that particular theory as to what sort of a thing a sentence-meaning is.

[69] See Neale [1], Lambert [1]. Also, Gareth Evans advocates Russell's theory in *The Varieties of Reference*. See Evans [1] chapter 2 especially pp. 51-60. Simon Blackburn defends the Theory of Descriptions in *Spreading the Word.*

[70] In 1990 Stephen Neale published a brilliant defense of Russell's theory (Neale [1], chapters 5 and 6.) But the topic of D-questions and D-imperatives and D-optatives – of D-sentences that are not assertoric -- is not mentioned even once, let alone discussed. And in 1984 Simon Blackburn wrote a brilliant exposition and defense of Russell's theory, which masterfully showed how a Russellian could deal with the fact that e.g. 'the man in the corner is bald' might be *de re* about a certain man (which *prima facie* favors the view that 'the man in the corner' is functioning as a referring expression and not *pace* Russell as a kind of quantifier) and with the fact that, if we heard 'the king of France is bald' we would be unlikely to reply 'that's false' (which *prima facie* suggests that 'the King of France is bald' does not, *pace* Russell, *assert* that there is a king of France.) But Blackburn completely ignores the question whether the theory of descriptions applies to definite descriptions occurring in non-assertoric contexts. Kent Bach published an extended defense of the Theory of Descriptions, in which he ably challenged a number of traditional attacks on it. But he does not even a single time discuss D-questions or D-imperatives. In the anthologies on definite descriptions that I have read through, the only D-sentences discussed are D-assertions. (See e.g. Ostertag [1].) There is not a single mention of – let alone discussion of – the question whether Russell's theory is viable so far as definite descriptions occurring in non-assertoric contexts are concerned. For some reason, contemporary discussions of the viability of Russell's theory tend to concern anaphora (See e.g. Neale [1]) and the apparently referential function performed by (some uses of) definite descriptions,

e.g. ‘the man in the corner is an economist’, ‘the table is covered with books.’ (See Kripke [1], Grice [1], Bach [1].)

[71] The F-subscript stands for ‘formal’.

[72] In this connection, I am using ‘good’ to mean ‘reasonable’ (not necessarily ‘valid’ in the technical sense – for (b) is probably not technically valid), and ‘bad’ to mean ‘unreasonable’.

[73] This is, ironically, Neale’s own example of an ambiguous sentence.

[74] Notice that, if you disambiguate a sentence *un*charitably, there is a feeling of strangeness or discomfort. Consider the following sequence: ‘Bob has no self-control. He can’t strop drinking. He can’t stop smoking. **He can’t stop gambling**.’ Now force yourself to disambiguate the boldfaced sentence *un*charitably (i.e. read it to mean *he can’t prevent people in general from gambling*). As long as you force yourself to read it that way, there is a feeling of discomfort. But once you disambiguate charitably – once you read it to mean *he cannot resist his own compulsion to gamble* – the sentence falls into place, and your feeling of discomfort vanishes. But now consider the following sequence: ‘Bob has no self-control. He can’t strop drinking. He can’t stop smoking. **He can’t prevent gambling**.’ The last sentence is *not* ambiguous between a meaning that is appropriate and one that is inappropriate. It is *incorrigibly* or *irremediably* out of place; and this is manifested by an ineradicable feeling of oddness (as opposed to the eradicable oddness borne by ‘He can’t stop gambling’ in (b).

[75] I think that this point goes back to Grice (though I do not know whether he makes it in exactly this form). See Grice [2].

[76] One might object: “Suppose A said ‘Does your *one* neighbor drink?’ Wouldn’t that normalize his response? “ I have two points to make in response. We’ve already seen (section II) several reasons not to think that this tactic is available to the Russellian. For example, A’s question seems to be a *strict* use of a definite description, if there ever was one (it is ‘proper’). But even if we leave that aside, the objector’s point is doubtful. To my ear the following dialogue *still* sounds markedly strange even if we replace A’s response with ‘does your *one* neighbor drink?’ And the most obvious diagnosis of why it sounds strange is the one we’ve given all along: that B is required, not to *presuppose*, but to *assert* the uniqueness and existence of A’s neighbor, and that by using a definite description – even one with ‘one’ added – he has really only presupposed this.

[77] Dummett makes some remarks that might be seen as responsive to this question. See Dummett [1], p. 316.

[78] This point was made to me, in response to an earlier draft of the present paper, by Nathan Salmon.

[79] (ii) is also ill-formed, but for some reason, the use of ‘for some x’ instead of ‘there is some x such that’ seems to have an ameliorative effect; it to *some* extent – though, to my ear, by no means completely – disguises the fact that (ii) is ill-formed.

[80] And also (ii).

[81] It will be noticed that here I am talking as though force operators are applied *not* to bits of verbiage but to the *meanings* (propositions or fragments thereof) borne by bits of verbiage. This is indeed how I wish to be understood. Force-operators *do not apply to sentences*; they *apply to propositions.* Any (well-formed) sentence, by virtue of its *grammatical* properties, *necessarily* embodies a force. ‘Bob went to the store’ *already* embodies a force; and so do ‘did Bob go to the store’ and ‘make

Bob go to the store.' So force-operators cannot significantly be applied to *sentences* – doing so would be redundant. However *propositions* are initially force-less. The sentences 'did Bob go the store' and Bob went to the store' and 'would that Bob went to the store' all encode the same *proposition* (*Bob went to the store*). But by virtue of their *grammatical* properties, these sentence embody the application of *different* force-operators to that same proposition. Basically the application of a force operator is expressed in natural language by such things as word order (in English the order of subject and verb is inverted in questions) and inflection (the tone of a question tends to be a rising one, whereas that of an assertion tends not to do this); in short, the application of a force operator is, in natural language, 'written into' the grammar of a sentence (and is not usually indicated by the application of a single character). To sum up, it is *propositions* that are assigned forces, not sentences. Sentences come into existence *already having* force: a well-formed sentence is *ipso facto* something that has a force. Not so of a proposition.

[82] This point was made to me, in private discussion, by both Francis Dauer and Nathan Salmon.

[83] I would conjecture that this is not a peculiarity of the English language; that it is universal.

[84] Notice I am not saying that the proposition in question would be *Bob went to the store and did Bob buy some cheese.* I am taking it for granted that 'Bob bought some cheese' and 'did Bob buy some cheese' are associated with the same *proposition* (*Bob bought some cheese*) but have different *forces.* I think it's a good idea *not* to try to assimilate force to propositional content: we want to be able to say that one can assert, ask, command one and the same proposition (at different times). If we reduce forces to components of propositions, then a given proposition will be capable *only* of being (say) asserted, not of being asked or commanded; and this is probably not the case.

[85] See the last footnote.

[86] Of course, oftentimes in writing, we encounter sentences that begin with 'And': e.g. 'people are mean. And they are greedy too.' But we shouldn't be misled by orthography here. 'And they are greedy' cannot *by itself* be assigned a force. Obviously what is assigned a force is that sentence in conjunction with the preceding one. To make this clear, suppose someone came up to you out of the blue and said: '*And* people are greedy'. You could obviously make sense of 'people are greedy', but not of the 'and'. You would assume, I think, that *either* you were hearing the second conjunct of some conjunction *or* that you were hearing something that wasn't quite well-formed.

[87] Technically, there is the option of saying that some sentences are, as wholes, to be assigned 'mixed forces' –are e.g. both questions and assertions – but (ii) that those sentences contain no well-formed assertoric part and/or no well-formed interrogative part. So 'Bob went to the store and did Bob buy some cheese' is *irreducibly* a question-assertion: as a whole, it is a question-assertion, but it cannot be neatly divided into a part that is a question and a part that is an assertion. It would take us too far afield to argued against this. But surely I am not alone in feeling that such a proposal stretches the limits of the intelligible. In any case, it is *ad hoc.*

[88] It may initially seem strange to say that, in saying (iii), I have commanded you to make a *conditional proposition* be true when we consider some related locutions, e.g.:

(v) I want the following to be the case: if Bob comes home, then you will tell him to call me.

There is nothing strange about (v) (except that it is long-winded and pedantic). Consider a close-cousin of (v):

(vi) I order that the following be the case: if Bob comes home, then you will tell him to call me.

(vi) is not strange. And (vi) is the assertoric analogue of:

(vii) Make the following be the case: if Bob comes home, then you will tell him to call me.

(vii) of course is just a long-winded, but otherwise normal, way of saying

(viii) Make the case that [if Bob comes home, then you will tell him to call me]

or, more simply,

(ix) If Bob comes home, then tell him to call me.

[89] See Grice [1], p. 270.

[90] This objection was made by an anonymous reviewer at *Principia.*

[91] Suppose I say: "I wish to know whether the king of France went to the store; therefore there is something I wish to know." At the level of literal meaning, those words express a valid argument. But the following does not express a valid argument, at least at the level of literal meaning: "Did the king of France go to the store? Therefore there is something I wish to know." That is nonsense. It is hard to see how replacing a synonymous sentence with a synonymous sentence could turn good sense into nonsense. So surely "I wish to know whether the king of France went to the store" is not synonymous with "did the king of France go the store?"

[92] This objection is due to an anonymous reviewer at *Principia.*

[93] These are the exact words of the reviewer mentioned in the last footnote.

[94] Quine 1982 (23).

[95] Clarence Lewis (1929) argued compellingly that *every* statement about the physical world – even those that seem to have no counter-factual component – are in fact implicitly counterfactual. When you say that something is steel, you are saying something about how it *would* behave *if* it were subjected to certain conditions. I am inclined to agree with Lewis. There is no way to purge our discourse of counter-factual elements. So we have to find a way to make counterfactuals work.

[96] Lewis 1973 (1).

[97] When they are true.

[98] Though perhaps not in all his views, e.g. his modal realism. But the truth is that Lewis' modal realism is, I think, common sense through to its logical conclusion. It is usually thought that Lewis' modal realism – his belief that possible worlds are *real* worlds that are not identical with this one – is a monstrosity created by a logician with his head in the clouds. The truth is, in my view, that the ordinary person thinks of counterfactuals in factualist terms: as expressing relations among concrete things – states of affairs – with the qualification that these concrete things don't exist here. But unless one is willing to take a completely non-factualist view – is willing to see counterfactuals as something like disguised probability statements – then it is (I think) impossible to avoid modal realism. For any state of affairs is per force a *concrete* state of affairs: something that exists in space and time. A 'non-concrete' state of affairs must be an obscure way of saying 'proposition' or 'concept'. And , of course, a concrete thing can exist only in a concrete world, not a *merely* possible one. For a 'merely possible concrete world' is an oxymoron.

Lewis' modal realism is, in this respect, like Wittgenstein's picture theory. The picture theory has always been reviled. But if one takes seriously the idea that a proposition's being true consists in its 'corresponding' with the facts – and how could one not take such a view? – then, of course, one must say what this 'correspondence' consists in. And – although the picture theory is probably false – the picture theory is the only *non-trivial* conception, thus far produced, of what that correspondence consists in. I think the picture theory delineates a conceit that is operative in virtually everyone's thinking about propositions: only, before Wittgenstein came along, it was never stated clearly, so the falsity of it was never seen. Inevitably, Wittgenstein was attacked. But really *we* should be attacked for endorsing the picture-theory (if, in fact, it is wrong). Similarly, although *Lewis* is attacked for his modal realism, he is, I think, simply saying in a logically developed way what *everybody* thinks in an inchoate, confused way. So if modal realism is wrong (which, I argue, it is: at least in connection with counter-factuals), it is not Lewis who should be attacked, but us.

[99] Blackburn 1984 (215) mentions this move and makes some incisive remarks about it.

[100] Lewis (1973: 77-78) himself gives criticisms of M that are very similar to the ones I give.

[101] This example was borrowed from Carnap, who makes this very point.

[102] Isaac Asimov wrote an interesting short story about this. (I can't remember the title.) The basic idea was that, if the law of gravity were suspended, the results would not be the pedestrian and comical ones we would expect (people floating upwards, and so on), but amounted to a total annihilation of reality as we know it.

[103] Of course, we can run counter-nomic statements ('if the inverse-square law didn't hold, but otherwise the initial conditions in experiment E were the same…') But, first of all, those sorts of counterfactuals are quite rare; people seldom assert them. Second, everything we said about 'ordinary' counterfactuals – i.e. counterfactuals where the laws are kept the same, but the events are varied – applies tenfold to counter-nomic statements. If a given physical law didn't hold, then the probability of *anything* being even remotely like anything in our world, on any level, would be essentially zero. A variation in *laws* would be much harder to contain than a variation in some fact. More exactly, both such things are impossible to contain, but the former are more obviously impossible to contain.

[104] I should mention that, in *Experience and Prediction*, Reichenbach *does* say that there are degrees of truth. In fact, he says in some places that there is no such thing as 100% truth. I think that Reichenbach is confusing epistemological with metaphysical (or semantic) issues. But leaving that aside, I highly doubt that even Reichenbach would characterize (*) as *true*. It isn't even clear if he'd characterize it as 73% true.

[105] Keynes' work on probability is based on this point, which he very clearly states at the beginning of his treatise. See Keynes 1921 (sections 2-3).

[106] I also allow that *all* probability statements are purely epistemic.

[107] We must make this assumption. It isn't worth considering the case where stupidity affects what answers the person gives. I am assuming that counter-factuals are *objectively* true or false – they are not mere expressions of how we feel. (Some, amazingly, have argued that they are expressions of feeling. I think that *ultimately* – despite his numerous disclaimers - Blackburn has essentially this view. See Blackburn 1984 (chapter 6) and Blackburn: 1993. I am arguing that *correct* counter-factuals are statements of the form $P_1..P_n$ gives such and such credibility to P_{n+1}. Obviously I want to illustrate this point by discussing people who assign the *right* probability of P_{n+1}, given $P_1…P_n$, not the ones who do it the wrong way.

[108] An atomic proposition is one that is not a generalization of some kind and is not built out of other propositions.

[109] And this is precisely the answer given by Hintikka, Carnap, and countless others.

[110] I myself hold a version of this view. It has to be refined a bit to avoid some set-theoretical paradoxes. But the basic idea is unquestionably correct. If P entails Q then thre is obviously an *internal* relationship between P and Q: a relationship that couldn't possibly hold between two *distinct* things. So some kind of containment relationship must hold between them. In my book *Conceptual Atomism and the Computational Theory of Mind* (unpublished) I defend, at great length, this very conception of propositionality. And I show why, contrary to first appearances, that conception *is* consistent with the fact that entailments are often informative.

[111] This example has been used many times. Goodman and Hempel use it – to illustrate this very point.

[112] The only philosopher I know of who denies (L) is Bas van Fraassen. But van Fraassen's conception of natural law and of counter-factuals does not involve the notion of possible worlds at all.

[113] Not one I agree with.

[114] The obvious response is this:

Suppose that Bob and Bob* have the exact same beliefs concerning what *has happened*, but Bob knows about laws that Bob* doesn't know about. For example, Bob knows that metal expands when heated, Bob* doesn't know this, but their respective knowledge of what has *happened* – of what states of affairs have obtained – is the same. In that case, Bob and Bob* can *rationally* assign different probabilities to, say, the possibility of some metal's expanding upon being heated.

Fine. But remember that whatever probability can reasonably be attached to a putative law of nature, the *sole* basis for that evaluation must lie in what has happened: *ultimately* all probability-assessments flow from what *has* happened and, indeed, from what is known. Of course, Bob may believe in some law L *not* because he has personally witnessed states of affairs that confirm it, but because some scientist told him about it or because he read about it in a book written by an authority. But in this case, Bob's 'knowledge' is obviously parasitic on other people's knowledge of singular states of affairs. *In the end*, what gives probability to a position is the totality of known *singular* states of affairs. So the way to respond to the objector is to say: if we leave aside parasitic knowledge (like Bob's knowledge of the laws of physics) *or*, alternatively, if we take into account the *singular* knowledge had by the communities to which Bob and Bob* belong, and from which their knowledge of laws is derivative, *then* there cannot be a divergence in the probability-assessments of these two people *unless* one of them is being irrational or, in any case, unless one of them is evaluating the data less competently than the other.

[115] Kripke 1980.

[116] At any rate, it used to be epistemically possible. Perhaps now it is definitive of the concept of *water* that, if something isn't H_2O, it doesn't fall under that concept.

[117] See Blanshard 1961 (20-22) for an insightful discussion of the probative value of one's 'feeling of freedom'.

[118] Grice 1990 and Kripke 1991.

[119] This point is borrowed from Kripke 1991.

[120] Just as, in my view, 'probably' would be ambiguous if it had both epistemic and non-epistemic meanings. That is why I think, even in statements like 'the probability the electron will jump is .8' the 'probably' is really epistemic. The electron will either jump or it won't. It could be that there's no law, and no hidden variable, that 'makes' it jump now as opposed to later. It could be that it just jumps. Let us suppose that to be the case. Where, in that picture, is there room for *non*-epistemic probability? I see room for epistemic probability: in the past, 80% of cases like this one have involved the electron jumping, and the other's have not. So the current situation gives an 80% probability to the proposition *the electron jumps.* But what about 'objective' probability? What is it for something *other* than a proposition to probabilify something other than a proposition? Do the circumstances involving electrons *make* the electron jump 80% of the time but not the other 20% of the time. That would be incoherent. By assumption there is no hidden variable; nothing is making anything happen in the one set of cases that isn't also operative in the other. So does the 'objective' probability in this case mean that the state of the universe prior to the electron jumping is 80% responsible for its jumping? That is even more blatantly incoherent. When we have an irreducibly indeterministic law – a law that doles out probabilities (of less than 100%) that we can't exchange for deterministic laws, no matter how much we come to know – we are still dealing with purely epistemic probabilities: the only difference is that, in this sort of case, an increase in knowledge won't change the probabilities, whereas when we are dealing with *reducibly* probabilistic laws (e.g. *75% of people shot in the chest die*), we can, with increases in knowledge, exchange that law for ones that dole out probabilities closer to (and sometimes coincident with) 100%.

[121] Actually, the word 'if' *is* ambiguous: sometimes it means 'whether' ('I don't know if Bill is coming back') and sometimes it denotes a relationship of conditionality or truth-dependence ('if Smith eats too much, he won't lose weight'). But if we confine ourselves to the latter kind of 'if' – which, obviously, we are in the present context – then it is very hard to see how 'if' could actually have different linguistic meanings in 'if arithmetic were complete, then robots could think' (counter-logical 'if') and 'if Bob had been nicer to his department head, he would have gotten tenure' (counter-factual 'if').

[122] For the record, I think that sentence-tokens *do* refer. But I haven't space enough to make a case for this here.

[123] But curiously, their compelling criticism of the Slingshot seems to have had little effect on discourse about the Slingshot. For example, it is nowhere mentioned in Neale's 2001 book on the Slingshot; nor is it mentioned in Neale's 1994 article on it.

[124] See Kaplan 1989. I believe that this point is found, in some form, in Strawson 1951.

Let me explain the "very roughly" qualification. A token of "that I am tired" is not true or false. If I say, simply, "that I am tired", I have not said anything true or false, even though the proposition *denoted* by my words is true or false. On the other hand, a token of "I am tired" *is* true or false. So there is a semantic difference between a token of "I am tired" and a token of "that I am tired", even though, plainly, those expressions are closely related in meaning. So the semantic content that a token of "I am tired" is assigned by the corresponding type is different from the semantic content had by "that I am tired". After all, a token of "I am tired" is true or false, whereas a token of "that I am tired" is neither. Thus the type "I am tired" does more than assign propositions to its tokens; it seems to assign propositions *plus* assertoric force-operators. So it is not entirely accurate to say that the semantic content of "I am tired" is a function that assigns propositions to its tokens: the function in question

assigns propositions *and* it also assigns a force to the propositions. But in this context, it will help discussion if we speak approximately, and say that the semantic content of the type "I am tired" is a function that assigns propositions to its tokens. This slight inaccuracy won't vitiate any of the points argued for here. And it can be corrected by mentally replaced the term "proposition" with "proposition plus assertoric force".

[125] See Barwise and Perry 1999, and also McCawley 1981. This point pervades linguistic literature.

[126] Keep in mind the caveat stated in the first footnote. The meaning of the type "I am tired" is a function that assigns propositions *plus forces* to its tokens; that function doesn't *merely* assign propositions to its tokens – though, of course, it does do that.

[127] Given the right circumstances, one could use any term, no matter what its literal meaning, to refer to anything. I could sarcastically use the term "Genghis Kahn" to refer to Bob. Such a usage is non-literal. I can use "Plato" in a *literal* way to refer to Plato. Obviously this fact is fact is essentially connected to the fact that "Plato" refers to Plato.

[128] Give or take some subtleties relating to force. Strictly speaking, the semantic content of a token of "I am tired" is *in part* a proposition. But that semantic content also has assertoric force. That is why a token of "am I tired" won't have the same exact semantic content as a token of "I am tired", even though both may encode the very same proposition

[129] The special cases I have in mind are analytic propositions like "Socrates is identical with Socrates". It isn't even clear that a grasp of the semantic content of *these* sentences necessarily involves a grasp of their truth. Even if it does, that has nothing to do with semantics; it has to do with epistemology.

[130] There is an obvious – and erroneous -- way to deal with this. One can say:

"You say that 'presumably, the referent of a referring term is at least part of its semantic content'. But your presumption is wrong. The referent of a referring term is *not* a part of its semantic content. "The inventor of bifocals" has the same semantic in both w and w*, even though it refers to different things in those worlds. This is perfectly acceptable. What an expression refers to simply doesn't have anything to do with what its semantic content is."

The position is a non-starter. "Plato" refers to Plato. Its having that referent is, to say the least, an integral part of its semantics. If "Plato" referred to Aristotle it would have a very different semantics. So it really isn't feasible to deny that what a term refers to is an integral part of its semantic content.

Frege actually *did* hold that, even if "Plato" referred to Aristotle, it wouldn't necessarily have a different semantic content from what it actually has. For the semantic content of "Plato", in Frege's view, is a *sense*; and so long as it "Plato" keeps the same sense, it keeps the same semantic content: it doesn't matter whether its referent changes. So a Fregean might deny our claim that reference is essential to semantic content.

Later on, we'll see that, if Frege *were* right about "Plato", then "Plato" would actually be a quantifier, and not an expression that referred to some individual. But our defense of this will have to wait.

But right now we can make a couple of points that render Frege's position innocuous, at least in this context. First of all, given what Kripke said, it is pretty clear that Frege is wrong about "Plato". Kripke's work also makes it clear that, if "Plato" had a different referent, it *would* have a different

semantic content. Putnam's work also makes this clear. If "water" referred to xyz, as opposed to H_2O, it *would* have a very different semantic content.

So, to say the least, it is unnatural to deny that, if a term refers, its reference is integral to its semantic content.

Conceivably, one could counter by saying:

"Granted, the referent of a proper name, like 'Plato', is part of its semantic content. But this is not the case with other referring terms (e.g. 'you', 'the inventor of bifocals'). In some cases, reference is essential to semantic content; in others, it is not."

But this is an extremely tortured and arbitrary view. It is clear that where some referring terms (e.g. proper names) are concerned, reference and semantic contribution are inextricably linked. If we say that this is not the case where other referring terms are concerned, we are in a position like that of someone who says "atomic structure is essential to some elements, but quite inessential to others". So it is surely best *not* to say "the referent of a referring term is no part of its semantic content". To put it in positive terms, if an expression refers to something, its referring to that thing is an integral part of its semantics. A corollary is that if two expressions refer to different things, then they have different semantics.

So we *still* have a conundrum. If two expressions refer to different things, then they have different semantic contents. At the same time, it seems that if "the inventor of bifocals" is a singular term, then it can refer to different things in different worlds, while keeping the same semantics.

[131] Strawson 1951.

[132] I actually think this does not give quite the right rule. Where definite descriptions are concerned, the context of utterance is defined by much more than just the world in which the tokening occurred. But the niceties are not important here. What is important here is that if "the inventor of bifocals" is taken as a non-quantifier, then what a token of it refers to is semantically dependent on the context of utterance, in *some* sense of the word "context".

[133] It should be made clear that the foregoing discussion has nothing to do with the *actual* semantics of definite descriptions. Some hold that "the inventor of bifocals" is not a singular term; it is a quantifier, and "…the inventor of bifocals…" means (roughly[133]): *something x uniquely invented bifocals and…x…*So "the inventor of bifocals snored" means: *something x uniquely invented bifocals and x snored.*

Even if this view is right, it in no way confutes our principle that, if a term refers, one must know what it refers to know its semantic content. Quite the contrary. Let us define "ze phi" as follows. *If* there some x such that x is a unique phi, then "ze phi" refers to x. If there is no such x, then "ze phi" is abortive. Everything we just said about "the phi" applies to "ze phi". Suppose that, in w and w*, "ze inventor of bifocals" has the same semantics. But in w*, it refers to Franklin, whereas in w* it refers to Jefferson. Here we seem to have a case where a single expression has a semantics but variable reference. But, for reasons we've seen, we don't want to say this; and there is an easy way to avoid saying this. The *type* "ze inventor of bifocals" has the same semantics in both w and w*. But the semantic content of the type is a function that assigns referents to its tokens. The type itself doesn't refer to anything. So our principle continues to be upheld. (For our principle is: *if* an expression refers to something, one must know to what it refers if one is to know its semantics.) *Tokens* of "ze inventor of bifocals" are what refer. The semantic content of such a token is what the function described a moment ago assigns to it; so the semantic content of such a token is its referent. Thus, to know the semantic content of a token of "ze inventor of bifocals", one must know what that

referent is.

[134] This is how Richard Montague, along with many other contemporary semanticists, think of "and", "not", and other connectives.

[135] This line of thought is obviously not an innocuous one, and perhaps requires development. I would like to outline how that development would go. The sentence:

(sn) "snow is not white"

is equivalent to:

(sn*) not* <snow is white>,

where "not*" ascribes falsity to the proposition encoded in the sentence-token falling in its scope. So the "not" in (sn) can be seen as a second-level property – on that ascribes falsity to the proposition encoded in the sentence-token falling in its scope. (Equivalently, it can be seen as encoding a function that assigns truth to a sentence-token exactly if the proposition encoded in that token is false.) The reason that "snow is white" occurs truth-functionally in (sn) and (sn*) is *not* that it contributes a truth-value; it is that a truth-value (falsity) is *ascribed* to the proposition it encodes. So we see, once again, that for a sentence-token to occur truth-functionally is *not* for it to contribute a truth-value; it is for a truth-value to be *ascribed* to the proposition that it encodes.

The case is not significantly different when we're dealing with atomic sentences. But here we have to proceed very cautiously. Obviously the truth of

(sw) "snow is white"

depends on whether the proposition encoded in that sentence (namely *that snow is white*) has the property of truth. We've seen that, when it occurs in the scope of "and" or "not", "snow is white" doesn't contribute a truth-value. It would be highly unnatural to suppose that "snow is white" makes one semantic contribution in (sg) and a different one in (sw). It very much seems to make the same contribution in both. We've seen that, in (sg), "snow is white" does not contribute a truth-value.

Given all of this, here is what I would propose. In (sw), "snow is white" isn't flanked by any verbiage. So it is occurring in what is sometimes called a "null" context. Null contexts, as linguists know, are often semantically pregnant. (They are null in a phonetic, not necessarily a semantic sense.) In English, when a sentence-token occurs in a null-context, truth is to be ascribed to the proposition encoded in that sentence-token. The null context is thus a phonetically invisible way of ascribing truth to the proposition encoded in the sentence. But here the relevant point is that what ascribes truth – what contributes the truth-value *true* – is, strictly speaking, the null context, not the token "snow is white" *per se*.

Here is another way of looking at it. Whether the sentence-token:

(st) "that snow is white is true".

is true is a function of whether *that snow is white* is true. As we saw, this is equivalent to saying that (st) attributes truth to that proposition. Obviously (sw) is a function of whether *that snow is white* is true. So (sw) attributes truth to that proposition. Thus, (st) attributes truth to *that snow is white* and so does (sw); and neither does anything besides that. (st) is true exactly if *that snow is white* is true, and the same holds of (sw). So the differences between (st) and (sw) do not lie in what they say. Those differences appear to have nothing to do with content and to be entirely phonetic. In the case of (st), the ascription of truth is phonetically realized. In the case of (sw), it is not phonetically realized. Given this,

it seems reasonable to say that, in the case of (sw), the *absence* of phonetically realized neighboring material performs the very function that, in the case of (st), is performed by phonetically realized material.

[136] I am innocuously modifying his argument

[137] I am slightly adapting Church's argument.

[138] Davidson 1967. The technique used in his argument is really due to Church 1943.

[139] See Davidson 1984, Neale 2001.

[140] This line of thought is found in Montague's groundbreaking work.

[141] See Barwise and Perry 1983 (399). They make this exact point.

[142] Frege 1892a.

[143] See the introduction to Whitehead and Russell 1912.

[144] Strawson 1950.

[145] The same is true with Gödel 1944, Quine 1953, and Davidson 1967.

[146] Compare Barwise and Perry 1999 (44-45): "[T]here is…something profoundly misleading about the traditional concern over entailments between sentences. For one thing, their concern obscures the efficiency of language, and it is this efficiency that is central to meaning. For example the *sentence* SOCRATES IS SPEAKING does not follow from the *sentences* EVERY PHILOSOPHER IS SPEAKING [and] SOCRATES IS A PHILOSOPHER even though this argument as the same "logical form" (on most accounts of logical form) as [ALL MEN ARE MORTAL, SOCRATES IS A MAN, THEREFORE SOCRATES IS MORTAL]. In the first place, there is the matter of tense. At the very least the three sentences would have to be said at more or less the same time for the argument to be valid…A rather startling consequence of this is that there can be no syntactic counterpart, of the kind traditionally sought in proof theory and theories of logical form, to the semantic theory of consequence. For consequence is simply not a relation between purely syntactic elements."

[147] For example, Church 1943, Gödel 1944, Quine 1953, Davidson 1967 all fail for that reason (and for others).

[148] Arguably, some will say that in some languages (e.g. Russian), one doesn't say "Bob is tired"; one says simply "Bob tired". Thus, it might be proposed, these languages are rife with non-indexical sentence-types.

That is very shallow reasoning. The way that the present-tense marker is expressed in such a language *is* by dropping the verb "to be" in contexts where it is understood to be the main connective.

[149] Quine makes a similar point in *Methods of Logic*.

[150] I'm adapting Church's argument a bit.

[151] Davidson 1967. The technique used in his argument is really due to Church 1943.

[152] See Barwise and Perry 1990: 399. They make this exact point.

[153] 1892a.

[154] 1956.

[155] Davidson 1990/1967.

[156] Quine 1953, 1960. 1966.

[157] Quine 1953, 1960, 1966.

[158] Frege actually gives two quite distinct arguments for the thesis that sentences refer to truth-values. For reasons of space, I will consider only the second (the slingshot). If cogent, the first argument – the one I don't give – shows that sentences *do* refer to truth-values. If cogent, the second one – the famous one (the "slingshot") – only shows that *if* sentences refer to anything, it is to truth-values.

[159] Strawson 1951.

[160] Russell 1990/1905. Whitehead and Russell 1997/1912 (introduction).

[161] Frege 1990/1892a.

[162] 1990/1951.

[163] I actually think this does not give quite the right rule. Where definite descriptions are concerned, the context of utterance is defined by much more than just the world in which the tokening occurred. But the niceties are not important here. What is important here is that if "the inventor of bifocals" is taken as a non-quantifier, then what a token of it refers to is semantically dependent on the context of utterance, in *some* sense of the word "context".

[164] Kripke 1971 makes similar points. He says that "the number that comes after four" is *de facto* rigid, meaning that, for reasons lying outside of semantics, it denotes the same thing in all worlds. And he says that "Nixon" is *de jure* rigid, meaning that for reasons lying *within* semantics it denotes the same thing in all worlds.

[165] Quine 1951. Putnam 1962.

[166] In some extremely broad and tenuous sense of "conceivably".

[167] I say "typically" because it has been argued (with considerable plausibility, in my view) that one cannot be acquainted with the contents of consciousness without knowing everything about them (leaving aside, of course, extrinsic facts about them).

[168] With some qualifications, I believe this point can be extended to the distinction between intension and extension. The word "whale", we are told, has both intension and extension. The intension is a property or function (from objects to the truth-values). The extension is the class of whales (Shamu, Willy, Moby Dick…)

But there is no one expression that has both intension and extension. Tokens of "whale" have extension; the expression-type "whale" have intension. Nothing has both.

We don't want to say that the expression-type "whales" denotes *different* classes in different worlds, while having the same semantics. That is perilously close to absurdity.

What we say is: *tokens* of "whales" denote different classes in different worlds; the *type* has an invariant meaning: it pairs off a token of "whale" in a world W with a class C iff C is the class containing al and only whales in W.

So intensions are the semantics of expression-*types.* Extensions are the semantic content of expression-tokens.

So what has "intension" is entirely different entity from what has "extension".

Having said this, the parallel between "the inventor of bifocals" and "whales" is not perfect.

Consider the sentence "whales are mammals". This is capable of very different readings. (Whether we are dealing with a semantic distinction, or merely a pragmatic one, is a question I leave open.) It is capable of both "referential" and "attributive" readings. The attributive reading is this:

(A) *if* a thing is a whale, then it is a mammal.

The referential reading is:

(R) Shamu and Willy and Moby Dick…are mammals.

The referential reading says *of* those things that are whales that they are mammals. The referential reading is to the effect that the things that are *actually* whales are mammals; so the referential reading makes the proposition in question turn on the mammal-hood of Willy, Shamu, and so on. It is object-involving with respect to those creatures.

The same is true (mutatis mutandis) if *all* plural nouns. They can be read attributively, as quantifiers, or they can be read referentially, as terms that pick out the things that *in fact* have the relevant property.

If "whales' is a quantifier, then one can understand a token of "whales are mammals" without knowing exactly which things are whales. (This is the reading that Frege (1884) favored.) If "whales" is what it appears to – a term that refers to the actual whales that exist – then one could *not* understand a token of "whales are mammals" without knowing exactly which things are whales.

This would *seem* to favor the attributive analysis of "whales". But such a conclusion would be premature. (Similar conclusions favor the view that Kaplan's "Newman 1" is not a referring term. But it is a referring term.)

But, in this context, the relevant point is that, if *anything* has for its extension the class of those things that are whales in this world, that thing is not the expression-type "whale": rather *tokens* of that type have that extension. If "whales" is a quantifier, then nothing – not the type "whale" and not tokens of it – has that class for its extension. Either way, there is never any one thing that has both intension and extension. If anything has that extension, it is *tokens* of "whales" (or "men" or "swans"), and not the type "whale". And if anything has that intension, it is the *type* "whale" (or "men" or "swans"), not tokens thereof.

Granted, matters are more complicated than I've made them out to be. For there *different* attributive readings of, for example, 'whales are mammals", and *different* referential readings of that sentence. Sometimes when a person says "x's are y's", he means that *some* x's are y's ("people kill other people"); sometimes he means that all x's are *in fact* y's ("all antelopes are in Brazil"); sometimes he means that not all x's are y's, but there is some kind of tendency for x's to be y's ("people are mean"); sometimes he means that, as a matter of necessity, nothing can be an x without being a y ("all pieces of metal conduct electricity", "all triangles are planar figures". I myself do not think that plural nouns have different meanings corresponding to all these different readings. The significant semantic question is: is "whales" ambiguous between referential and attributive readings? It is, I think, safe to assume that the various sub-readings within those two categories are generated by pragmatics. In any case, the principle we've been arguing for holds up no matter how the actual semantics turns out; there is never any one entity that has both intension and extension.

A historical point might be made here. Frege said that “whales are mammals” doesn’t say anything about whales. People balk at this. But this reasoning is good. According to him, that sentence means: *if* a thing is a whale, it is a mammal. It doesn’t say that *this or that particular* thing is a mammal. So if we consider the whales that exist – Shamu, Willy…-- “whales are mammals” does not say of any one of them that it is a mammal. Therefore “whales are mammals” doesn’t say anything about any whale.

So Frege’s view is plausible. But it has the counter-intuitive consequence that “whales are mammals” doesn’t say anything about any actual whale.

It should be pointed out that there is a reading of “whales are mammals” on which it *does* make a statement about whales. That is the referential reading. In my view, the referential reading is probably the one that is closest to its literal meaning. On the referential reading, “whales are mammals” says of those things that are actually whales that they are mammals.

But the referential reading has the counter-intuitive consequence that one doesn’t know what (a token of) “whales are mammals” means unless one knows exactly which things are whales and which aren’t. And no human being does know that.

[169] Davidson (1990/1967).

[170] See Neale (2001: 4-8).

[171] There are probably type-theoretic reasons why it is wrong to speak of *the* class of objects x such that (x=x). In that case, let Omni be the class of all such objects of a given logical type.

[172] It follows that there is one linguistically expressible fact.

[173] It is a delicate question whether that pair comprises the *sentences* or propositions. For reasons I will argue later, I think it is pretty clear that it comprises propositions. But in the present context, this is not important.

[174] This term is borrowed from Neale (2001). It is short for Principle of Substitution of Logical Equivalents.

[175] See Barwise and Perry (1990: 398-400):

“Church gives a form of the slingshot….Church considers these sentences:

(5) Sir Walter Scott is the author of Waverly.
(6) Sir Walter Scott is the man who wrote the twenty-nine Waverly novels altogether.
(7) The number, such that Sir Walter Scott is the man that wrote that many *Waverly* novels altogether, is twenty-nine.
(8) The number of counties in Utah is twenty-nine.

Church argues that we as go from each sentence to the next, what the sentences denote is the same. But the first and last seem to have nothing in common except their truth-value; he says, “elaboration of examples of this kind leads us quickly to the conclusion, at least plausible, that all sentences have the same denotation.”

Sentence (6) results from (5) by replacement of one description by another, where both descriptions describe the same person, Scott. Sentence (*) results from (&) by a similar move, with the descriptions describing th same number, twenty-nine. The step form (6) to (7), however, is of a different sort. Church says that (6), though perhaps, not synonymous with (&), “is t least so nearly so as to ensure having the same denotation.”

The argument is like an ambiguous figure or an Escher drawing. *If you are aware of situations*, you have to keep shifting perspective to let the argument trick you. From one perspective the first and last steps re fine but the middle step is all wrong. From a second perspective the middle step is reasonably good but he fist and last steps are completely unfounded.

Intuitively, situations are complexes of objects and relations (under which we shall from now on subsume properties). Given this conception, the role of the parts of the sentence to identify objects and relations out of which the complex is constructed, Let us hold this in mind, and go through the steps of the ar4gument from both perspectives.

To get from (5) to (6), we need to suppose that the great difference between "the author of Waverly" and "th mean who wrote twenty-nine Waverly novels altogether" makes absolutely no difference to the situations described by the two sentences, that is, that the contribution these two descriptions make to the situation is just to identify Scott. Thus, the first perspective is the one where all four definite descriptions are interpreted by the objects they happen to describe, Scott in two cases, the number twenty-nine in the other two.

But from this perspective, the step from (6) to (7) does not work at all. Recall we are attempting to show that all four sentences must designate the same situations. But from our current perspective [that according to which sentences designate situations], (6) designates a situation whose only constituent object is Scott, whereas, (7) designates one whose only constituent object is the number twenty-nine.

Now let us focus on the step from (6) to (7). If we attempt to see these sentences as designating the same situation, then it must be that of Scott's having written exactly twenty-nine *Waverly* novels altogether. *To see them as designating this situation*, however, we must pay close attention to the properties involved in the definite descriptions. This is not unreasonable. *If you take situations seriously*, it quite natural to distinguish two ways in which descriptions might contribute to the sentences. The need is for the material to build t a situation, a complex of objects and properties, and why should the description not contribute the individual described or the properties involved in the description of the situation?

Thus, the second perspective, which is suggested anyway if we take situations seriously, is the one where "the author of Waverly" is not interpreted simply as Scott, but contributes the complex of objects and properties it mentions to the situation the sentence describes. But this perspective is absolutely fatal to the step from (5) to (6) or from (7) to (8). The descriptions of Scott, and those of twenty-nine even more so, contribute radically different objects and properties to the overall situation. Sentence (7) designates twenty-nine's being the number of Waverly novels Scott wrote, but sentence (8) designates twenty-nine's being the number of counties in Utah: distinct situations if ever there were distinct situations.

So we see that from the first perspective, the one Frege would have us take, wherein the reference of the description all that is specific is obliterated, the first and last steps of the argument are fine but eh middle is all wrong. From the second perspective, more in line with Russell's theory of descriptions, the middles step comes of better but the first and last steps are wrong. Under neither reading are we compelled to accept the argument."

For right now, let us focus only the very last sentence of this passage: "Under neither reading are we compelled to accept the [slingshot]."

The readings in question are, basically, the one where the definite description contributes only its referent and the one where it contributes a complex of properties: the referential and the attributive.

Read referentially, the argument fails. Read attributively the argument fails.

This is the real point that Barwise and Perry are making. And it is correct.

But in their exposition of this point, they keep on adverting to their situation-semantics; they make it *seem* as though the argument they are presenting requires us to sign off on the idea that sentences contribute *situations*.

Their argument does *not* really require this. But their exposition makes us think otherwise.

And this, presumably, accounts for criticisms like the following (Olson 1998: 96-97):

"Thus, according to Barwise and Perry, we are free to interpret our descriptions any way we like, so long as we choose one interpretation and stick to it, instead of changing in the middle of an argument.

This is a persuasive diagnosis, and the accompanying prescription certainly removes the symptoms. But *it relies quite explicitly upon the structuralist criterion*, taking for granted that where there are different constituents there must be different facts."

Olson then goes on to say how, according to some viewpoints, one and the same fact can be analyzed into different constituents. Therefore, the "structuralist criterion" is questionable. Therefore, Olson concludes, what Barwise and Perry say against the slingshot is questionable.

Olson is wrong. What Barwise and Perry say *about the slingshot* doesn't require that one adopt a "structuralist" or any other conception of facts or propositions or situations or meanings. Their attack on the slingshot is totally neutral on these matters.

Their thesis that sentences stand for situations is another matter. But that is a different thing from their attack on the slingshot.

Barwise and Perry (1999: 23-26) give a shorter version of their argument against the Slingshot. The problem is that the argument they give there is far worse than the argument just quoted. In Barwise and Perry (1999) they simply assume that logically equivalent sentences may *differ* in reference. This obviously makes their argument less than compelling.

Neale (2001, 1995) devoted an entire book, and a very long article, to the Slingshot. (The article was absorbed into the book practically verbatim.) Astonishingly, he says essentially nothing about Barwise and Perry (1990) or Barwise and Perry (1999). He mentions Barwise and Perry (1990), but somehow doesn't register its importance or even its relevance, except to some minor technical problems.

Perry (2000: 290-294) restates the view just quoted from Barwise and Perry (1990/1983). But Perry's (2000) restatement is even more opaque than the 1990 version.

[176] Russell 1921.

[177] It is first put forth, to my knowledge, by Kripke 1972. See also Evans 1990/1973.

[178] Contemporary Russellians would say that the "there is" is subject to context-based restrictions. This point is critical in discussions of whether Russell's' theory is correct. But it doesn't have any special importance here. So for the sake of brevity, I am skipping over it.

[179] Donnellan 1990/1966.

[180] Strawson (1990) talks about "predicative" uses of definite descriptions. He says (rightly, no doubt) that in "Caesar was the greatest general of all time", the definite description attributes a property to Caesar; it is *not* functioning as a referring term.

[181] As it happens, for reasons I don't give here, I think that, at the level of semantics definite descriptions *are* referring terms, without exception. Even when they are being *used* predicatively or attributively, their semantic function is merely to refer.

[182] Church 1956.

[183] Frege 1892a.

[184] Mill 1991 (10).

[185] Church 1956.

[186] I say “arguably” because, quite frankly, I really don’t think that either Frege or Mill had a sufficiently developed conception of the distinction between semantics and pragmatics – between what is literally meant by an expression and what is pragmatically imparted by it – to warrant saying that they had definite views as to the *manner* in which “the phi” contributes descriptive content. Granted, Frege (1879) did make some important points about the distinction between the “cognitive content” of sentences and their “coloring” or “tone”. And this has a certain isomorphism to the distinction between semantics and pragmatics (though, I believe, it is not quite the same distinction). But Frege was, with all due respect, quite foggy on the various different ways in which “the phi” might contribute its sense.

[187] Here, of course, people will bring up Donnellan’s point that, even when “the man drinking the martini” is functioning as a *referring* expression (not a quantifier), “the man drinking the martini smokes” makes a true statement (or, better, can be *used* to make a true statement) even if there is no man drinking a martini. And this fact might call into question the accuracy of my statement that (REF) captures the semantics of “the phi” for referentialists. This is well-trodden territory. It is pretty clear that *if* “the man drinking a martini” is a referring expression, *then* at the level of literal meaning “the man drinking the martini smokes” is abortive (neither true nor false) if there is no one drinking a martini. Granted, that sentence may say something true. But that is very easily explained as a pragmatic epiphenomenon. And it would be an egregious violation of Grice’s “modified Occam’s razor” (never multiply meanings without necessity) to say that *semantically* “the man drinking the martini” actually referred to (say) the female transvestite drinking the ginger ale.

[188] I am idealizing away from the pragmatic epiphenomena mentioned in the last footnote. I don’t think they have any relevance here.

[189] Russell 1990/1905, Neale 1990.

[190] *Pragmatically* “Socrates” probably is associated with descriptive content – though that descriptive content will tend to vary. If somebody points to a man wearing a toga and says “that is Socrates” (or “ ‘Socrates’ is the name of that individual”), my introduction to the semantics of “Socrates” is via the informational content of a specific sense-perception. (The same thing is true *mutatis mutandis* if my introduction to the semantics of that term is via some statement like: “Socrates is the guy who is the protagonist in most of Plato’s dialogues.”) Inevitably, such facts will affect what “Socrates snores” imparts to me, and will (perhaps) give a *cognitive* content in excess of the bare singular proposition *Socrates snores* ,even though it is that singular proposition, and nothing besides, that is semantically encoded in that sentence. Nothing Kripke (1972) says rules out this kind of pragmatic association.

[191] Sticklers for Russell-exegesis may balk at my use here of his distinction. But it seems to me that, leaving aside niceties, there can be no doubt that a distinction more or less like the one Russell had in mind has a place here. Indeed, Russell *himself* makes this exact distinction, and uses the same terminology to describe that distinction, in his paper “Knowledge by Acquaintance and Knowledge by Description”. (He is not, perhaps, employing this distinction in connection with quite the same class of sentences I’m discussing. but the similarities are unmistakable.)

[192] See Kripke 1972, Evans 1990/1973.

[193] At least not to *most* such terms: for most are formed by putting a “that” in front of a sentence expressing that proposition; and any such term has constituent structure. It is, of course, possible to

invent a term having *no* constituent structure that denotes some proposition. We could say: let "A" refer to the proposition *snow is white*. Conceivably, though I doubt it, the causal theory might apply to terms like A.

[194] I put RING in brackets because, in L, it is not the English word "ring", but the *actual* ringing of some particular doorbell that occurs in that sentence-token.

[195] Causal theorists might allow that different causal relations will do.

[196] I put RING in brackets because, in L, it is not the English word "ring", but the *actual* ringing of some particular doorbell that occurs in that sentence-token.

[197] This extra semantic rule is phenomenally ungainly and useless. But it does apparently qualify as a genuine semantic rule.

[198] I would guess.

[199] I forget exactly what that word is. An Urdu-speaking friend of mine told it to me. But it is not "Socrates".

[200] Blackburn (1984) argued that Tarski's "The Concept of Truth in Formal Languages" disjunctives the concept of truth: there is trueinEnglish, trueinUrdu. This is perhaps not quite on point. But I though it worth mentioning.

[201] It is, I believe, a completely dispensable part. Consider the following semantic rule. If there some person x who uniquely was a protagonist in most of Plato's dialogues, then let "SOC" refer to x. If there is no such person, then let "SOC" refer to nothing.

Obviously "SOC" refers to Socrates every bit as much "Socrates" does. But it seems that one's being able to use "SOC" to refer to Socrates doesn't involve having any particular causal connection to Socrates.

The causal theorist is guilty of the same fallacy as the descriptivist. Both confuse reference-fixing with meaning-giving. The descriptivist says: "Socrates" *means* "the unique hemlock drinking philosopher". He says that to refer *is* to describe. Descriptions are certainly *exploited* by semantic rules, as we just saw. We can invent rules that assign Gödel to "Gödel" *by way of descriptions*. (An example is: if there is some x such that x uniquely proved the incompleteness of arithmetic, then let "Gödel" refer to x. Here a description is being *exploited* to assign a referent to a word. But the reference relation – not here, not anyway – is *identical* with the describing relation. Similarly, causal relations can be exploited by semantic rules. We can say "if there is some entity x such that uniquely stands in R [some causal relation] to tokens of 'Igaguk', then let any such token refer to x." Here a causal relation is being *exploited* by a semantic rule assigning a referent to a word – just as a description can be. But no causal relation – not here, not anywhere – is ever *identical* with the reference relation. It were absurd to think otherwise.

[202] I am borrowing a felicitous expression from Kent Bach.

[203] Frege 1997/1891, Frege 1997/1892b.

[204] I myself have no doubt that quantifiers *are* referring terms and that they refer to second-order properties. But, strictly speaking, the properties in question are properties, not of propositions or of open-propositions, but of *sequences* of such things. The quantifier "for all x" refers to a property had by certain *sequences* of open-propositions. "For all x, x is mortal, if x is a man" says: The sequence <x is a man, x is mortal> is such that anything instantiating the first member instantiates the second.

So, in that case, "for all x" denotes the property *is such that anything instantiating the first member instantiates the second.* And this is a property of sequences of some kind.

[205] In any case, there is nothing tendentious about it in this dialectical context. Perhaps in discussions of modality or counterfactuality it might be a tendentious assumption.

[206] The asterices are meant to be quasi-quotes. My computer doesn't do quasi-quotes.

[207] See Barwise and Perry 1999 (6).

[208] When "that" functions this way, linguists say it is operating as a "complementizer". Many languages permit the *dropping* of the "complementizer", at least in some contexts. I would like to know, but do not know, whether this is true of all languages. I would conjecture that it is. And this fact (if it is a fact) supports (albeit weakly) my point – the author's point, not the hostile interlocutor's point – that "Bob smokes" and "that Bob smokes" refer to the same thing.

[209] Of course, "that grass is green" is not, at least not phonetically, identical with "grass is green". And (though I think this is implausible) one might suppose that dropping the "that" induces a change in reference: previously a proposition was referred to; now a truth-value is referred to.

But then we are just stuck, once again, with the view that "grass is green" refers to a truth-value in (p) and either refers to nothing, or refers to a proposition, in (n). And this is just what we want to avoid saying.

[210] We could imagine a language L just like English except that, in L, one gave assertoric force to "that grass is green" by saying it in a certain one of voice or, perhaps, by modulating one's voice in a certain way. So L-translation of "grass is green" would be "that grass is green", said in a certain way. Speakers of L wouldn't dare say that "that grass is green", said assertorically, and "that grass is green", said non-assertorically, would differ in reference.)

[211] But, as Russell once said, grammar is a *prima facie* good guide to logical form. Logical form and grammatical form pull apart. But a good first clue to a sentence's logical form is its grammatical form.

[212] "And" doesn't always have this function. If you say "Russell and Whitehead wrote the *Principia Mathematica*" you are not saying "Russell wrote the *Principia Mathematica* and Whitehead wrote the *Principia Mathematica.*" You are saying, it seems, that some entity composed of *both* Russell and Whitehead wrote that work.

So "and" sometimes joins *individuals*; it doesn't always join propositions. But it always joins *something.* So we do not, arguably, have a case of sheer ambiguity.

To determine whether "and" *actually* has a different one meaning in "Russell and Whitehead wrote the *Principia Mathematica*" and another one in "snow is white and grass is green", we'd have to get very clear on what "and" is doing in the latter. Some clarity will be attained, I hope, in the present work.

[213] Frege did not, to my knowledge, talk about the operator "possibly". But what we said about other operators, in particular epistemic ones, makes it clear what he would have said or, at any rate, was compelled to say, about "possibly".

[214] It is assumed that I am being sincere, that I know the meanings of those words, and that I am using them in a literal way.

[215] Again, I am confining my attention to sincere, literal, linguistically competent utterances.

[216] Actually, I think the real satisfaction-condition is: *x is a unique contextually salient inventor of bifocals.* But this doesn't matter in this context.

[217] Stalnaker 1991.

[218] Church 1943 (299).

[219] 1953.

[220] 1944 (129-130).

[221] On the Fregean view, "1+1=2" and "triangles have three sides" do indeed co-refer. Indeed, on the Fregean view, they co-refer, as a matter of logical necessity.

At first this seems to be an advantage over my view.

It is not.

When two *sentences* are said to be "logically equivalent", S-equivalence is not what is meant. If you ask a competent person, "what is it for two sentences to be logically equivalent?", he absolutely will not say: "it is for the open proposition that a thing must satisfy to be picked out by the one expression to entail, and be entailed by, the open proposition that a thing must satisfy to be picked out by the other." What he *will* say is: it is for the proposition meant by the one to entail, and be entailed by, the proposition meant by the other. So the kind of "logical equivalence" that people have in mind when they talk of logically equivalent *sentences* is not the kind of "logical equivalence" that people have in mind when they talk about expressions like "Plato" and "the unique x such that 1+1=2 and x=Plato". When it said that two *sentences* are "logically equivalent", S-equivalence is not meant; so the kind of "logical equivalence" that guarantees co-reference is *not* what is meant. Some completely different kind of logical equivalence is meant. And this other kind of logical equivalence has very little to do with co-reference. So to think that "1+1=2" and "triangles have three sides" must co-refer because they are "logically equivalent" is to make the false and confused assumption that those two sentences are logically equivalent in the sense in which "Plato" and "the unique x such that 1+1=2 and x=Plato" are logically equivalent. We have two very different notions of logical equivalence here, not one. The one kind of logical equivalence involves co-reference. But that leaves it totally open whether the other kind involves co-reference.

Let us close the argument. It is often thought that 1+1=2" and "triangles have three sides" must co-refer. Frege thought this; so do all proponents of slingshot arguments. The reasoning behind this view is:

Those two sentences "logically equivalent"; and "logically equivalent expressions" – e.g. "Plato" and "the unique x such that 1+1=2 and x=Plato" – must co-refer.

But that view involves a complete failure to distinguish to kinds of "logical equivalence". And the view that a correct semantic theory must make logically equivalent sentences co-refer is based on that very confusion. So, therefore, is the view that Frege's analysis has an advantage over mine, in that his view makes "logically equivalent" sentences co-refer while mine does not. The idea that that is an advantage involves a confusion of two very different meanings of the term "logically equivalent".

[222] Davidson 1984/1965.

[223] Neale 2001, 1994.

[224] Chomsky 1965.

[225] Frege 1953/1883: x. Writing in a slightly different context, he also said: "[W]e ought always to keep before our eyes a complete proposition. Only in a proposition have the words really a meaning. It may be that mental pictures float before us all the while, but these need not correspond to the logical elements in the judgment. It is enough if the proposition taken as a whole has a sense; it is this that confers on its parts also their content." (*Ibid*: 71).

[226] Dummett 1973 (192).

[227] In my view, that major accomplishment of that work was not the reduction of numbers to sets. It was the recognition, and defense, of two truths: logical form and grammatical form diverge; and, what is related, statements that *seem* to be about individuals are in fact statements about *objects.* When you say "a man smoked", you are making a statement about a *concept*, not about some odd individual.

[228] Of course, this is not the only way to define "some man". It can also be treated as a referring term; it can be defined by saying that it refers to a function that assigns truth to a class C iff that class comprises a man. But, as I will now show, even if "some man" is treated in that way, it must *still* be defined contextually; for all referring terms are to be do defined.

[229] Of course, this was not Frege's view; Frege thought that "Plato" had a "sense" for its meaning. But here we are not concerned with what Frege thought "Plato" actually meant; we are concerned with how, according to Dummett, Frege would have *interpreted* the supposition that the meaning of "Plato" is Plato.

[230] See Neale (2001: 19-32) for a critique of truth-conditional semantics that is similar to the one I'm about to give.

[231] See Neale (2001: 23).

[232] See Neale (2001: 30).

[233] Quine 1966b (54).

[234] But, for our purposes, what is important is whether that analysis is correct, not whether Frege held it. So I will ignore the question whether Tughendat has correctly analyzed Frege.

Strictly speaking, what I am about to say isn't *quite* what Tughendat says. Tughendat talks not about the *meanings* of sentences, but about their truth-conditions. And Tughendat does not exactly talk about the contributions that words make to the *meanings* of sentences; rather, he talks about their contribution words to make to the *truth*-values of sentences – their "truth-value potential".

Tughendat's is writing an article about how to interpret Frege; he is not writing an article about how to analyze reference. That may justify his talking about truth-values instead of meanings. But what is relevant here is Tughendat's larger point: for a word x to refer to a thing y just *is* for x to contribute y to the semantic value of sentences in which x occurs. Tughendat says that the semantic value of a sentence is a truth-value.; I say it is a proposition. But we agree that reference is to be understood in terms of the concept of the semantic value of a sentence. And that is the relevant point here.

[235] Tughendat says that the "significance" of a sentence is its truth-value. I am going to take the "significance" of a sentence to be its meaning – to be the thing that would ordinarily be called its "significance". (I can know the "significance" of "Bush smokes" without knowing its truth-value.) This does, I grant, somewhat distort his argument. But it also helps to expose what is relevant and

significant in that argument. Also, Tughendat's thesis is that the meaning of a word is the contribution it makes to the "significance", i.e. the truth-value, of a whole sentence. But not all sentences have truth-values – there are questions and imperatives. And sentences are not always asserted; they can be negated, or occur in sub-ordinate clauses. Perhaps these facts can be reconciled with Tughendat's view that the meaning of a word is its contribution to the truth-value of a sentence. But it seems simpler to stick to the more obvious view that the meaning of a word is its contribution to the meaning of a sentence; and thus not have to make special adjustments for cases of non-asserted sentences.

[236] Tughendat talks not about the *meanings* of sentences, but about their truth-conditions. I don't know whether this is because that is how Frege talks or for some other reason. In any case, the difference between meanings and truth-conditions doesn't play an important in this discussion.

[237] See Kaplan 1989. I believe that this point is found, in some form, in Strawson 1951.

Let me explain the "very roughly" qualification. A token of "that I am tired" is not true or false. If I say, simply, "that I am tired", I have not said anything true or false, even though the proposition *denoted* by my words is true or false. On the other hand, a token of "I am tired" *is* true or false. So there is a semantic difference between a token of "I am tired" and a token of "that I am tired", even though, plainly, those expressions are closely related in meaning. So the semantic content that a token of "I am tired" is assigned by the corresponding type is different from the semantic content had by "that I am tired". After all, a token of "I am tired" is true or false, whereas a token of "that I am tired" is neither. Thus the type "I am tired" does more than assign propositions to its tokens; it seems to assign propositions *plus* assertoric force-operators. So it is not entirely accurate to say that the semantic content of "I am tired" is a function that assigns propositions to its tokens: the function in question assigns propositions *and* it also assigns a force to the propositions. But in this context, it will help discussion if we speak approximately, and say that the semantic content of the type "I am tired" is a function that assigns propositions to its tokens. This slight inaccuracy won't vitiate any of the points argued for here. And it can be corrected by mentally replaced the term "proposition" with "proposition plus assertoric force".

[238] See Barwise and Perry 1999, and also McCawley 1981. This point pervades linguistic literature.

[239] Keep in mind the caveat stated in the first footnote. The meaning of the type "I am tired" is a function that assigns propositions *plus forces* to its tokens; that function doesn't *merely* assign propositions to its tokens – though, of course, it does do that.

[240] Given the right circumstances, one could use any term, no matter what its literal meaning, to refer to anything. I could sarcastically use the term "Genghis Kahn" to refer to Bob. Such a usage is non-literal. I can use "Plato" in a *literal* way to refer to Plato. Obviously this fact is fact is essentially connected to the fact that "Plato" refers to Plato.

[241] Review of *Frege on Functions* by Max Black. Journal of Symbolic Logic 1956 (21): 201-202.

[242] Actually, I will argue that "John" picks out a third-level function. But we can ignore that here.

[243] Give or take some subtleties relating to force. Strictly speaking, the semantic content of a token of "I am tired" is *in part* a proposition. But that semantic content also has assertoric force. That is why a token of "am I tired" won't have the same exact semantic content as a token of "I am tired", even though both may encode the very same proposition

[244] The special cases I have in mind are analytic propositions like “Socrates is identical with Socrates”. It isn’t even clear that a grasp of the semantic content of *these* sentences necessarily involves a grasp of their truth. Even if it does, that has nothing to do with semantics; it has to do with epistemology.

[245] There is an obvious – and erroneous -- way to deal with this. One can say:

“You say that ‘presumably, the referent of a referring term is at least part of its semantic content’. But your presumption is wrong. The referent of a referring term is *not* a part of its semantic content. “The inventor of bifocals” has the same semantic in both w and w*, even though it refers to different things in those worlds. This is perfectly acceptable. What an expression refers to simply doesn’t have anything to do with what its semantic content is.”

The position is a non-starter. “Plato” refers to Plato. Its having that referent is, to say the least, an integral part of its semantics. If “Plato” referred to Aristotle it would have a very different semantics. So it really isn’t feasible to deny that what a term refers to is an integral part of its semantic content.

Frege actually *did* hold that, even if “Plato” referred to Aristotle, it wouldn’t necessarily have a different semantic content from what it actually has. For the semantic content of “Plato”, in Frege’s view, is a *sense*; and so long as it “Plato” keeps the same sense, it keeps the same semantic content: it doesn’t matter whether its referent changes. So a Fregean might deny our claim that reference is essential to semantic content.

Later on, we’ll see that, if Frege *were* right about “Plato”, then “Plato” would actually be a quantifier, and not an expression that referred to some individual. But our defense of this will have to wait.

But right now we can make a couple of points that render Frege’s position innocuous, at least in this context. First of all, given what Kripke said, it is pretty clear that Frege is wrong about “Plato”. Kripke’s work also makes it clear that, if “Plato” had a different referent, it *would* have a different semantic content. Putnam’s work also makes this clear. If “water” referred to xyz, as opposed to H_2O, it *would* have a very different semantic content.

So, to say the least, it is unnatural to deny that, if a term refers, its reference is integral to its semantic content.

Conceivably, one could counter by saying:

“Granted, the referent of a proper name, like ‘Plato’, is part of its semantic content. But this is not the case with other referring terms (e.g. ‘you’, ‘the inventor of bifocals’). In some cases, reference is essential to semantic content; in others, it is not.”

But this is an extremely tortured and arbitrary view. It is clear that where some referring terms (e.g. proper names) are concerned, reference and semantic contribution are inextricably linked. If we say that this is not the case where other referring terms are concerned, we are in a position like that of someone who says “atomic structure is essential to some elements, but quite inessential to others”. So it is surely best *not* to say “the referent of a referring term is no part of its semantic content”. To put it in positive terms, if an expression refers to something, its referring to that thing is an integral part of its semantics. A corollary is that if two expressions refer to different things, then they have different semantics.

So we *still* have a conundrum. If two expressions refer to different things, then they have different semantic contents. At the same time, it seems that if “the inventor of bifocals” is a singular term,

then it can refer to different things in different worlds, while keeping the same semantics.

[246] Strawson 1951.

[247] I actually think this does not give quite the right rule. Where definite descriptions are concerned, the context of utterance is defined by much more than just the world in which the tokening occurred. But the niceties are not important here. What is important here is that if "the inventor of bifocals" is taken as a non-quantifier, then what a token of it refers to is semantically dependent on the context of utterance, in *some* sense of the word "context".

[248] It should be made clear that the foregoing discussion has nothing to do with the *actual* semantics of definite descriptions. Some hold that "the inventor of bifocals" is not a singular term; it is a quantifier, and "…the inventor of bifocals…" means (roughly [248]): *something x uniquely invented bifocals and…x…*So "the inventor of bifocals snored" means: *something x uniquely invented bifocals and x snored.*

Even if this view is right, it in no way confutes our principle that, if a term refers, one must know what it refers to know its semantic content. Quite the contrary. Let us define "ze phi" as follows. *If* there some x such that x is a unique phi, then "ze phi" refers to x. If there is no such x, then "ze phi" is abortive. Everything we just said about "the phi" applies to "ze phi". Suppose that, in w and w*, "ze inventor of bifocals" has the same semantics. But in w*, it refers to Franklin, whereas in w* it refers to Jefferson. Here we seem to have a case where a single expression has a semantics but variable reference. But, for reasons we've seen, we don't want to say this; and there is an easy way to avoid saying this. The *type* "ze inventor of bifocals" has the same semantics in both w and w*. But the semantic content of the type is a function that assigns referents to its tokens. The type itself doesn't refer to anything. So our principle continues to be upheld. (For our principle is: *if* an expression refers to something, one must know to what it refers if one is to know its semantics.) *Tokens* of "ze inventor of bifocals" are what refer. The semantic content of such a token is what the function described a moment ago assigns to it; so the semantic content of such a token is its referent. Thus, to know the semantic content of a token of "ze inventor of bifocals", one must know what that referent is.

[249] This is how Richard Montague, along with many other contemporary semanticists, think of "and", "not", and other connectives.

[250] This line of thought is obviously not an innocuous one, and perhaps requires development. I would like to outline how that development would go. The sentence:

(sn) "snow is not white"

is equivalent to:

(sn*) not* <snow is white>,

where "not*" ascribes falsity to the proposition encoded in the sentence-token falling in its scope. So the "not" in (sn) can be seen as a second-level property – on that ascribes falsity to the proposition encoded in the sentence-token falling in its scope. (Equivalently, it can be seen as encoding a function that assigns truth to a sentence-token exactly if the proposition encoded in that token is false.) The reason that "snow is white" occurs truth-functionally in (sn) and (sn*) is *not* that it contributes a truth-value; it is that a truth-value (falsity) is *ascribed* to the proposition it encodes. So we see, once again, that for a sentence-token to occur truth-functionally is *not* for it to contribute a truth-value; it is for a truth-value to be *ascribed* to the proposition that it encodes.

The case is not significantly different when we're dealing with atomic sentences. But here we have to proceed very cautiously. Obviously the truth of

(sw) "snow is white"

depends on whether the proposition encoded in that sentence (namely *that snow is white*) has the property of truth. We've seen that, when it occurs in the scope of "and" or "not", "snow is white" doesn't contribute a truth-value. It would be highly unnatural to suppose that "snow is white" makes one semantic contribution in (sg) and a different one in (sw). It very much seems to make the same contribution in both. We've seen that, in (sg), "snow is white" does not contribute a truth-value.

Given all of this, here is what I would propose. In (sw), "snow is white" isn't flanked by any verbiage. So it is occurring in what is sometimes called a "null" context. Null contexts, as linguists know, are often semantically pregnant. (They are null in a phonetic, not necessarily a semantic sense.) In English, when a sentence-token occurs in a null-context, truth is to be ascribed to the proposition encoded in that sentence-token. The null context is thus a phonetically invisible way of ascribing truth to the proposition encoded in the sentence. But here the relevant point is that what ascribes truth – what contributes the truth-value *true* – is, strictly speaking, the null context, not the token "snow is white" *per se*.

Here is another way of looking at it. Whether the sentence-token:

(st) "that snow is white is true".

is true is a function of whether *that snow is white* is true. As we saw, this is equivalent to saying that (st) attributes truth to that proposition. Obviously (sw) is a function of whether *that snow is white* is true. So (sw) attributes truth to that proposition. Thus, (st) attributes truth to *that snow is white* and so does (sw); and neither does anything besides that. (st) is true exactly if *that snow is white* is true, and the same holds of (sw). So the differences between (st) and (sw) do not lie in what they say. Those differences appear to have nothing to do with content and to be entirely phonetic. In the case of (st), the ascription of truth is phonetically realized. In the case of (sw), it is not phonetically realized. Given this, it seems reasonable to say that, in the case of (sw), the *absence* of phonetically realized neighboring material performs the very function that, in the case of (st), is performed by phonetically realized material.

[251] I am innocuously modifying his argument

[252] Davidson 1967. The technique used in his argument is really due to Church 1943.

[253] See Davidson 1984, Neale 2001.

[254] This line of thought is found in Montague's groundbreaking work.

[255] See Barwise and Perry 1983 (399). They make this exact point.

[256] Frege 1892a.

[257] See the introduction to Whitehead and Russell 1912.

[258] Strawson 1950.

[259] See Kaplan 1989.

[260] The asterices are meant to be quasi-quotes. My computer doesn't do quasi-quotes.

[261] The same is true with Gödel 1944, Quine 1953, and Davidson 1967.

[262] Compare Barwise and Perry 1999 (44-45): "[T]here is…something profoundly misleading about the traditional concern over entailments between sentences. For one thing, their concern obscures the efficiency of language, and it is this efficiency that is central to meaning. For example the *sentence* SOCRATES IS SPEAKING does not follow from the *sentences* EVERY PHILOSOPHER IS SPEAKING [and] SOCRATES IS A PHILOSOPHER even though this argument as the same "logical form" (on most accounts of logical form) as [ALL MEN ARE MORTAL, SOCRATES IS A MAN, THEREFORE SOCRATES IS MORTAL]. In the first place, there is the matter of tense. At the very least the three sentences would have to be said at more or less the same time for the argument to be valid…A rather startling consequence of this is that there can be no syntactic counterpart, of the kind traditionally sought in proof theory and theories of logical form, to the semantic theory of consequence. For consequence is simply not a relation between purely syntactic elements."

[263] For example, Church 1943, Gödel 1944, Quine 1953, Davidson 1967 all fail for that reason (and for others).

[264] Quine 1953a.

[265] It is a delicate question whether that pair comprises the *sentences* or propositions. For reasons I will argue later, I think it is pretty clear that it comprises propositions. But in the present context, this is not important.

[266] This term is borrowed from Neale (2001). It is short for: Principle of Substitution of Logical Equivalents.

[267] Church1944 (129-130).

[268] On the Fregean view, "1+1=2" and "triangles have three sides" do indeed co-refer. Indeed, on the Fregean view, they co-refer, as a matter of logical necessity.

At first this seems to be an advantage over my view.

It is not.

When two *sentences* are said to be "logically equivalent", S-equivalence is not what is meant. If you ask a competent person, "what is it for two sentences to be logically equivalent?", he absolutely will not say: "it is for the open proposition that a thing must satisfy to be picked out by the one expression to entail, and be entailed by, the open proposition that a thing must satisfy to be picked out by the other." What he *will* say is: it is for the proposition meant by the one to entail, and be entailed by, the proposition meant by the other. So the kind of "logical equivalence" that people have in mind when they talk of logically equivalent *sentences* is not the kind of "logical equivalence" that people have in mind when they talk about expressions like "Plato" and "the unique x such that 1+1=2 and x=Plato". When it is said that two *sentences* are "logically equivalent", S-equivalence is not meant; so the kind of "logical equivalence" that guarantees co-reference is *not* what is meant. Some completely different kind of logical equivalence is meant. And this other kind of logical equivalence has very little to do with co-reference. So to think that "1+1=2" and "triangles have three sides" must co-refer because they are "logically equivalent" is to make the false and confused assumption that those two sentences are logically equivalent in the sense in which "Plato" and "the unique x such that 1+1=2 and x=Plato" are logically equivalent. We have two very different notions of logical equivalence here, not one. The one kind of logical equivalence involves co-reference. But that leaves it totally open whether the other kind

involves co-reference.

Let us close the argument. It is often thought that 1+1=2" and "triangles have three sides" must co-refer. Frege thought this; so do all proponents of Slingshot arguments. The reasoning behind this view is:

Those two sentences "logically equivalent"; and "logically equivalent expressions" – e.g. "Plato" and "the unique x such that 1+1=2 and x=Plato" – must co-refer.

But that view involves a complete failure to distinguish two kinds of "logical equivalence". And the view that a correct semantic theory must make logically equivalent sentences co-refer is based on that very confusion. So, therefore, is the view that Frege's analysis has an advantage over mine, in that his view makes "logically equivalent" sentences co-refer while mine does not. The idea that that is an advantage involves a confusion of two very different meanings of the term "logically equivalent".

[269] 1964 (77).

[270] See Dummett 1973 (192-193) for similar remarks.

[271] See Kripke 1972, Evans 1990/1973.

[272] See Kripke 1971, 1972.

[273] See Kripke 1971, 1972.

[274] I put RING in brackets because, in L, it is not the English word "ring", but the *actual* ringing of some particular doorbell that occurs in that sentence-token.

[275] See Dummett 1973, and Tughendat 1970, for similar remarks.

[276] I put RING in brackets because, in L, it is not the English word "ring", but the *actual* ringing of some particular doorbell that occurs in that sentence-token.

[277] This extra semantic rule is phenomenally ungainly and useless. But it does apparently qualify as a genuine semantic rule.

[278] Blackburn (1984) argued that Tarski's paper "The Concept of Truth in Formal Languages" disjunctives the concept of truth: there is trueinEnglish, trueinUrdu, and so on. This is perhaps not quite on point. But I though it worth mentioning.

[279] Fodor 1998.

[280] See Fodor 1990a and 1990b for an attempt to deal with what are essentially these problems. Fodor is dealing, not the causal theory of reference, but with the causal theory of *conception*: for x to be a mental representation of y simply *is* for x to have a certain causal relation to y. The causal theory of conception is exactly analogous to the causal theory of reference. The problems that face the one are exact analogues of the problems that face the other. Fodor's two papers, though ingenious, show how implausible those theories are.

[281] Kaplan 1989b (604-607) makes some remarks that are similar to these.

[282] Kaplan 1989b (604-607) makes some remarks that are similar to these.

[283] Before we move one, there is an objection, due to Nathan Salmon, that we must consider (I

quote):

"You say that 'grass is green' has the same meaning in both

(3) 'grass is green'

and also

(4). 'Little Timmy believes that grass is green.'

But, contrary to what you say, this is perhaps consistent with its referring to a truth-value in (3) but not in (4). Roughly, it may function as a Millian name of a proposition in (3), and as a description of a truth-value in (4). If I am right about this, then we can hold onto the idea that "grass is green" has the same meaning in both (3) and (4), while also holding that it refers to a truth-value in extensional contexts."

I'm not entirely sure I understand this objection. But what I think it means is this.

Consider some definite description, say "the teacher of Aristotle". This description can function either referentially or attributively. If you say "Little Timmy believes that the teacher of Aristotle was wise", you may be saying that Little Timmy believes something about *Plato* (referential reading) or you may be saying that Little Timmy believes that *whoever it was* that taught Plato was wise (attributive reading).

There are very good reasons for thinking that "the teacher of Aristotle" is not semantically ambiguous, and that its amenability to both referential and attributive readings is a matter of

[283]

pragmatics, not semantics.

The objector seems to be saying that "grass is green" is, in this respect, like a definite description. In (3) it refers to a truth-value (it refers to it by encoding descriptive information that singles it out). So in that context it is functioning "referentially". But in (4), "grass is green" refers to a proposition. So it is functioning "attributively".

But even though "grass is green" may *function* in both ways, it doesn't follow that it is semantically ambiguous. The ambiguity might be completely pragmatic, like that of "the teacher of Aristotle".

So maybe, at the level of semantics, "grass is green" always refers to a truth-value. That it sometimes *seems* not to do so is some kind of pragmatic epiphenomenon.

Considerable support for the objection lies in the fact that, in all likelihood, "that grass is green" *is* a definite description ("the proposition that grass is green").

Let us now evaluate Salmon's point. First of all, if "(that) grass is green" describes anything, it probably isn't a truth-value. "That grass is green" is, as we noted, nearly synonymous with "the [ze] proposition that grass is green". And the latter describes, and refers to, a proposition, not a truth-value. [283]

But let us waive this; for the sake of argument, let us suppose that "grass is green" does describe a truth-value.

Obviously "grass is green" has different truth-values in different worlds. (Even if it doesn't, just replace "grass is green" with a sentence that *does* thus vary in truth-value.) Let w be this world, and let w* be some world where English is spoken but grass is purple.

If the objector is right, then in w, "grass is green" describes the truth-value *true.* In w*, it describes the truth-value *false.*

Since English is spoken in w*, "grass is green" has the same semantics there that it has here in w.

Obviously the referent of an expression is at least *part* of its semantic content.

Therefore, in w*, "grass is green" doesn't have (quite) the same semantic content that it has here in w.

But obviously "grass is green" has exactly the same semantic content in both cases. A dilemma.

We've dealt with this sort of situation before. The way to deal with it is to say:

"Grass is green" has a two-dimensional semantic structure. In a world where grass is green, *tokens* of it refer to (by describing) the truth-value *true*. In a world where grass is not green, tokens of it refer to the truth-value *false*. So "grass is green" is a non-rigid designator; and whether a token of it picks out the true or the false is a function of the context (world) of utterance.

If this is right, then it is not "grass is green" *per se* that picks out a truth-value, but rather tokens of it. The expression "grass is green" *per se* wouldn't refer to anything; its semantic content would be a function of a certain kind.

A consequence is that the semantic content of a *token* of "grass is green" is a truth-value, while the semantic content of the expression-type is the just mentioned function.

Now we can see the problem with the objector's approach. If he is right, then one cannot know the semantic content of a token of "grass is green" unless one knows its truth-value in the context (world) of utterance (just as one cannot know the semantic content of an indexical unless one knows what it refers to in the context of utterance). In general, one cannot know the semantic content of a token of a sentence unless one knows its truth-value in the context (world) of utterance

But it is hard to believe that I don't know the semantic content of a token of a sentence-token if I don't know its truth-value. To understand a sentence-token – to understand a token of "that person is tired" – is to know what *would* have to be the case *if* it were true; it is not to know *whether* it is true. The most natural position, I think, is to say that one *can* understand sentence-tokens without knowing their truth-values; and, consequently, that such tokens do not refer to truth-values.

As we've seen, it is not easy to make the case that sentence-*types* refer to truth-values. For obviously the truth-value associated with the sentence-*type* "grass is green" varies from world to world, and possibly even from time to time, even though its semantics does not thus vary. So unless we are willing to take the heroic measure of saying that the referent of an expression is not part of its semantic content, we should distance ourselves from the idea that sentence-types refer to truth-values.

[284]

Everybody thinks it *strange* to suppose that "grass is green" means one thing on its own and a different thing in "John believes that grass is green". But most believe this supposition to be coherent.

There are a couple of reasons not to concede even this much to Frege's view.

a Kripkean reason

Frege says that certain operators cause sentences falling in their scope to undergo a shift in meaning. A consequence of this position is that repeated application of such an operator induces repeated shifts in meaning.

Frege's position is inconsistent with Kripke's discovery that, at the level of semantics, proper names do not have senses.

According to Frege,

(i) "Bertrand Arthur William Russell is Bertrand Arthur William Russell"

contributes a truth-value. In

(ii) "Fred believes that Bertrand Arthur William Russell is Bertrand Arthur William Russell".

"Bertrand Arthur William Russell is Bertrand Arthur William Russell " contributes a proposition: the proposition it expresses in (i).

When it occurs in

(iii) "Jerry believes that Fred believes that Bertrand Arthur William Russell is Bertrand Arthur William Russell",

"Bertrand Arthur William Russell is Bertrand Arthur William Russell " contributes a third thing: the proposition it *expresses*, not the one it refers to, in "Fred believes that Bertrand Arthur William Russell is Bertrand Arthur William Russell".

Frege's theory requires that, in (ii), "Bertrand Arthur William Russell is Bertrand Arthur William Russell" have both sense and reference. Its reference in (ii) is a proposition: *that Bertrand Arthur William Russell is Bertrand Arthur William Russell.* Its sense in (ii) is some proposition *through* which we grasp its referent.

We know from Kripke that proper names of individuals are not semantically associated with senses. Semantically "Bertrand Arthur William Russell" does not have a sense.

Frege's theory requires that "Bertrand Arthur William Russell" have a sense. Otherwise, that theory falsely assigns *no* meaning to oblique occurrences of "Bertrand Arthur William Russell". Since "Bertrand Arthur William Russell" doesn't have the required sense, Frege's theory fails.

Further, Frege's theory requires that "Bertrand Arthur William Russell" be *infinitely* ambiguous in two quite distinct ways. First, it requires that it mean one thing in (i), another thing in (ii), a third thing in (iii), and a fourth thing in

(iv) "Fred believes that Bob believes that Bertrand Arthur William Russell is Bertrand Arthur William Russell".

So Frege's theory makes "Bertrand Arthur William Russell" infinitely ambiguous in that way.

But we must remember that any object can be given by infinitely many different modes of presentation or senses. There are infinitely many descriptions that single out anything. So, if Frege's theory is right, then "Bertrand Arthur William Russell" is infinitely ambiguous in (iii), infinitely ambiguous in (iv), and so on.

But "Bertrand Arthur William Russell" isn't ambiguous at all, let alone doubly infinitely ambiguous.

Even if, by some remote chance, two people are named "Bertrand Arthur William Russell", so that it is ambiguous, it still wouldn't be ambiguous in the way required by Frege's theory.

The problems with Frege's theory are well documented. But what is not realized is that they bedevil any theory according to which non-truth-functionality involves meaning-shift.

Suppose you say that, in truth-functional contexts, sentences refer to truth-values. You must deal with the fact that they obviously *don't* refer to such things in other contexts. The most plausible way to deal with this is by supposing, as Frege did, that certain operators induce meaning-shifts. If you go this route, you must answer the question: What is the nature of that shift? What does the sentence refer to before the shift, and what does it refer to afterwards?

The most plausible answer is the one Frege gives: *after* the shift, the sentence refers to the thing that, prior to the shift, it had for its sense. But we've just seen the problems with this view.

A Davidsonian Argument

[284]

Donald Davidson put forth a clever argument that challenges Frege's analysis of indirect discourse. I do not know whether Davidson's argument is decisive. But it deserves to be mentioned.

[284] [284]

English is obviously a learnable language. Any learnable language has a finitely large primitive lexicon. If the Frege-Church line is correct, then "snow is white" means one thing on its own, a second thing in "Tim said that John believes that snow is white", and so on.

You cannot understand *all* of these sentences. But that has nothing to do with your *semantic* knowledge. It has to do with your ability to *exploit* your semantic knowledge. It has to do with matters relating to fatigue, limitations of memory, and general cognitive ability. If we considered a being that had exactly your semantic knowledge, but who had unlimited energy, logical prowess and so on, that being *would* be able to understand an occurrence of "snow is white" that was pre-fixed by 42,345 epistemic operators. If you could fully exploit your knowledge English semantics, you could assign meaning to such a sentence. Any semantic theory must be consistent with this fact. [284]

Let "snow is white$_1$" be synonymous with the sentence "snow is white" when it occurs on its own, or in an otherwise truth-functional context. Let "snow is white$_2$" be synonymous with "snow is white" when the latter occurs in the scope of a single epistemic operator. Let "snow is white$_3$" be synonymous with "snow is white" when the latter occurs in the scope of two epistemic operators. And so on.

Anyone who speaks English has the semantic resources to understand "snow is white$_{7,865}$". If Frege is right, for any i, your knowledge of English semantics comprises a knowledge of one semantic rule for "snow is white$_i$", and a different one for "snow is white$_{i+1}$". So your knowledge of English semantics comprises *infinitely* many semantic rules, corresponding to "snow is white".

By itself, this might seem acceptable. After all, English semantics *does* comprise infinitely many rules. But most of them are *derived*; and one can learn those on the basis of the primitive rules. If Frege's theory is right, then English comprises infinitely many *primitive* semantic rules.

Frege's theory, by implication, makes "snow" be *infinitely* ambiguous. There is "snow$_1$", "snow$_2$", "snow$_3$", and so on. According to that theory, "snow" is ambiguous between all these different words.

For each i, "snow$_i$" is a *primitive* entry in our lexicon. It has no constituent structure; no recursive rules give it meaning.

But, as we've seen, every English speaker *does* know, for any number n, what is meant by "snow" when it is prefixed by n epistemic operators. So Frege's analysis requires that each person who speaks a natural language know *infinitely* many primitive semantic rules. This is not acceptable. It makes English unlearnable. Besides that, it simply isn't plausible: English doesn't comprise infinitely many primitive semantic rules.

Granted, this argument is not airtight. One could make the following objection to it:

Frege does make "snow" infinitely ambiguous – there is "snow$_1$", "snow$_2$", and so on. But there is not a separate, primitive rule for each of these. There is one recursive rule, namely: "snow$_1$" means *snow*, and for i, "snow$_{i+1}$" refers to the sense of "snow$_i$ ".

Oddly, Davidson doesn't consider this.

In any case, this objection is easily met. For the sake of argument, let us suppose that, for each i, "snow$_i$" semantically encodes a sense. This supposition is false. But without it, Frege's argument fails. So let us grant it.

There is "no backwards road" from denotation to sense, Russell [284] once said. Any given thing can be given by infinitely many different senses. After all, any thing can be uniquely singled out by infinitely many different descriptions.

So, for each i, there are infinitely many possible senses that "snow$_i$ " could have. The recursive rule just described would have to choose some *one* sense. Any such choice would obviously be arbitrary; it

wouldn't correspond to any fact about the actual semantics of "snow". Further, the semantic rule described would have to select a single member from each one of an infinitely large number of infinitely large classes. It is, to my knowledge, an open question in set theory whether such a rule could exist.

I conclude that Davidson's argument has some force, and that attempts to meet it redound to the discredit of Frege's theory.

For reasons discussed earlier, Frege's view is not an anachronism. If we say that sentences refer to truth-values, then we must probably adopt Frege's view. We must say that "snow" is infinitely ambiguous, and that each of its disambiguations is given meaning by a primitive semantic rule. So, if we say that sentences refer to truth-values, we are saying, by implication, that English comprises infinitely many primitive semantic rules.

[285] David Lewis (1972) takes a similar position: see the next footnote.

[286] Lewis 1972 says that "does Plato snore?" and "Plato, snore!" *do* differ in propositional content. The former means: I want to know whether Plato snores. The second means: I command that you, Plato, snore. As Davidson 1984c (114-115) points out, this proposal is wildly implausible. And, as we will presently see, it is demonstrably false.

[287] Astonishingly, David Lewis has said that "does Plato snore?" and "I wish to know whether Plato snores" are semantically equivalent. But this simply isn't so, as we've just seen. Also, "I wish to know whether Plato snores; therefore there is something I wish to know" is, at the level of literal meaning, a valid argument. But "does Plato snore? Therefore there is something I wish to know" is, at the level of literal meaning, absolute nonsense. To say otherwise is to seriously overwork the concept of pragmatics.

[288] This line of thought pervades Austin's work. For example, see Austin 1962.

[289] Some historical background may put this into context. There are two very different conceptions of what linguistic meaning is. Some see sentence-meaning, and expression-meaning generally, in non-psychological terms. The meaning of "some" is a function from classes to truth-values; the meaning of "Plato" is an individual, or a function from contexts of utterance to that individual. The meaning of an expression is always a non-psychological entity, except in the special case where the expression in question denotes a mental content.

Let us refer to advocates of this view as "semantic objectivists".

Others see sentence-meaning in psychological terms. The meaning of a sentence-token is what the *speaker* means by it. Advocates of this view are sometimes referred to as "speech-act theorists". When you perform a speech-act – for example, when you make an assertion or ask a question – your words mean what *you* mean.

Of course, speech-act theorists grant that if someone says "I am cold" but means *I am hot*, his words don't mean what *he* means. I am simply outlining it. My statement of his position is meant only to be a bare approximation.

Speech-act theorists tend to embrace a semantic nihilism. They tend to think that the notion of *the* meaning of an expression-type is bankrupt: when we talk about *the* meaning of "snow is white", that is really just an oblique and misleading way of talking about what *people* mean when they produce those sounds or sounds relevantly similar to them; they think that talk of "meaning", in the platonic sense, is

[289]

really just a confused way of talking about "meaning", in the psychological sense.

I don't think that speech-act theory requires any kind of anti-Platonism. (In fact, I think it requires a kind of Platonism.) But, as a matter of historical fact, speech act theorists tend to be anti-Platonistic; they tend to say that meanings, in the non-psychological sense, don't exist.

So speech-act theorists tend to say, in effect, that semantics doesn't exist. Semantics is typically

thought of as an attempt to associate the platonic entities – the right functions, the right sets of worlds, the right properties – with expressions. The speech act theorist tends to reject this approach. So he rejects the discipline of semantics, at least as most semanticists conceive it.

For some reason, semantic-objectivists talk very little about force. But speech-act theorists talk a great deal about it. Since the only people who talk about force are the ones who reject the very notion of semantics, the impression has been created that force has no place in discussions of semantics.

This is plainly a mistake. Even those who have an extremely hard-line objectivist conception of meaning, such as the one advocated in this work, must deal with the concept of force. And they *can* deal with it. Psychology is no more needed to deal with the concept of force than it is needed to deal with the concept of variable-binding.

(I think psychology does have an important connection to semantics. But the connection is that mental entities *hook up* platonic entities to expressions, and thus *give* certain meanings to those expressions. The connection is *not* that mental entities *constitute* the meanings of expressions. I don't think there is any real opposition between semantic objectivism and speech-act theory. One is focusing on the psychological mechanisms that assign meanings to expressions. The other is focusing on the meanings themselves.)

[290] Dummett 1973 (316) and Geach (1965) say explicitly that force-operators are entirely "different" from other sentence-constituents, and that it is nonsensical to ask what they refer to or even what "sense" they have. The Dummett-Geach position, we will see, is based on an excessively narrow conception of reference – on a failure to see that any case of semantic contribution, any alteration of literal meaning, is easily and naturally construed as involving *reference.* But sometimes what is referred to is a function of some kind, and not an individual. Of course, changes (or additions) in sentence-force are changes in literal meaning; *pace* Austin (1962) they are obviously not just pragmatic or "perlocutionary" differences. (As we just saw, there *is* a semantic difference between "I wish to know whether Plato snores" and "does Plato snore?" In general, sentences with different forces can never have *quite* the same literal meaning, even though they can encode the same proposition.) It follows that force-operators *do* refer. In a moment, we will see what they refer to exactly.

[291] Dummett 1973 (315-316) says that force-operators don't denote anything and also that they don't have any identifiable meanings. (I don't know what he means by the latter.) Dummett provides, at best, minimal substantiation for these views. In any case, Dummett is simply wrong about this, as, I believe, we are seeing.

[292] Davidson 1984c discusses the nature of sentence-force. His main points are simply fallacious. To begin with, Davidson (119) says that "indicatives [i.e. sentences with assertoric force] we may well leave alone, since we have found no intelligible use for the assertion-sign" . So Davidson explicitly says that he will take the notion of *indicative* force as primitive.

But this is totally implausible. Obviously "snow is white" and "that snow is white" are semantically different (one is truth-evaluable, the other is not); also obviously, they both contain some constituent that encodes the proposition *that snow is white.* Obviously "snow is white" does *more* than just encode a constituent that encodes *that snow is white.* And, just as obviously, that extra ingredient is what gives "snow is white" assertoric force. So assertoric force is *not* primitive and it must be explained. There obviously *is* an explanation for the difference between "snow is white" and "that snow is white". And by taking assertoric force as primitive, Davidson blocks us from saying what that difference is.

This brings us to another of Davidson's points. According to Davidson, "go to the store" is semantically equivalent with: "The next utterance has imperatival force. You will go to the store." Here are Davidson's own words:

"In English we mark the non-indicative mood in various…ways, by changes in the verb, word-order, punctuation or intonation. We may think of non-indicative sentences, ten, as indicative sentence plus an expression that syntactically represents the appropriate transformation; call this expression the *mood-setter*. [Davidson's emphasis.]…[A] non-indicative sentence may be decomposed into an indicative sentence and a mood setter…"

A page later, Davidson gives an example of what he means:

"If we were to represent in linear form the utterance of, say, the imperative sentence "Put on your hat", it would come out s the utterance of a sentence like "my next utterance is imperatival in force", followed by an utterance of "you will put on your hat".

But there are many reasons, some of which we've already seen, why this proposal is far from the truth. First of all, it wrongly takes for granted that assertoric force is primitive. Second, it is viciously circular. Davidson is trying to define the concept of "imperatival force": but notice that he *uses* that concept in his analysis. He says that "put on your hat" has imperatival force because it really means: "my next utterance is imperatival in force: you will put on your hat". Finally, as we've seen, "put on your hat" is *not* semantically equivalent with "my next utterance is imperatival in force: you will put on your hat" or, for that matter, with any other indicative sentence. "Put on your hat" is not true or false; thus, at the level of literal meaning, it doesn't entail anything. But that is not the case with "my next utterance is imperatival in force: you will put on your hat" or with any other indicative sentence.

[293] Wittgenstein 1958, Dummett 1973.

[294] See Dummett 1973 (192-193) for similar remarks.

[295] 1999, 1990/1983.

[296] Quine 1966c.

[297] Quine 1966c.

[298] Quine 1966c.

[299] This is what linguists tend to say about it. See McCawley 1998.

[300] I believe Russell 1905 was the first person to make this important distinction.

[301] See Robinson 2002 (135).

[302] Russell 1903.

[303] Lyons 1977.

[304] Usually SSP's, as I am calling them, are referred to as "Sorites" paradoxes. But this terminology has become inextricably linked with a subtle, but not entirely insignificant, confusion. Consequently I will use the term "SSP" instead.

The confusion I am referring to is this. Consider the following argument:

(H1) Let S be a heap of rice. Removing one grain of rice from S won't make it stop being a heap. In general, given any heap of rice, removing one grain won't make it stop being a heap. So by induction, one grain of rice (or even zero grains) is a heap.

Of course, S could be a heap of anything – it doesn't have to be grains of rice.

A cousin of (H1) is:

(H2) Consider a single grain of rice R. By itself, R is not a heap of rice. Adding a single grain of rice to a non-heap will result in a non-heap. (It is not as though you turn a non-heap into a heap by adding a *single* grain of rice.) So if you add a trillion grains of rice to R, you still don't have a heap.

Of course, H2 isn't necessarily about grains of rice; it could be about discrete objects of many different kinds.

(H1) and (H2) are paradigm cases of so-called "Sorites"-type paradoxes. ("Sorites" comes from the Greek word "soros", which means *heap.*) Now (H1) and (H2) are certainly *similar* to what I am calling SSP's. But (H1) and (H2) actually have a more complicated structure. And, if I am not mistaken, the way to solve them is not the way to solve SSP's. A very approximate way of identifying the relevant difference is this. Let R be a grain of rice, or multiple grains of rice, lying on some surface. In order to generate a *heap* of rice out R, it is necessary not to just to add new grains of rice on the surface in question, but to add new grains *in a particular way* that surface. The newly added grains of rice have to occupy certain spatial positions relative to the grains of rice already on that surface (otherwise, one has a film of rice, or a coating of rice, not a heap of rice).

But nothing comparable to this is associated with the paradoxes I will be concerned with in the present paper – in what I am calling SSP's. In order to turn a slow-moving object into a fast moving object, it is necessary only to give it more speed; it isn't necessary to add that speed "in a certain way", whatever this might mean. As long as the speed has been added – as long as the object that *was* going one mph is now going 10,000,000 mph – we now have a fast-moving object. How that extra speed was added is irrelevant. In order to turn a non-wealthy person into a wealth person, it is necessary only to give him more money. It isn't necessary to give him money "in a certain way", whatever this might mean. As long as he has the extra dollars, he is wealthy. So SSP's different from Sorites-type paradoxes in that the operation of "adding" associated with the latter is subject to structural constraints not associated with the "adding" involved in SSP's.

[305] For a review of all the dominant solutions to the Sorites, see the paper entitled "Sorites" by Timothy Williamson and R.M. Sainsbury, which appears in *A Companion to the Philosophy of Language* (ed. Bob Hale and Crispin Wright). Williamson and Sainsbury do a splendid job of setting forth the dominant contemporary solutions – the use of alternative logics, the epistemic view (the idea, though we don't know it, that *there is* a sharp cut-off between being bald and not bald or between smart and not smart), the idea that there are degrees of truth, the supervaluationist approaches – and of showing why these solutions all have severe difficulties. The solution I will present is not among those considered by Williamson and Sainsbury.

See also ***Vagueness****: A* ***Reader***: Rosanna Keefe and Peter Smith (eds). There are several articles there that deal with SSP's (they are not distinguished from Sorites-type paradoxes), and each of them advocates an abandonment or a relaxation of some aspect of classical logic (usually bivalence).

[306] Unless, of course, the term is one which occurs only vacuously in a given argument, as the word "Bob" does in: Snow is white and Bob is tall; everything white reflects light and Bob smokes; therefore snow reflects light.

[307] See e.g. Dummett (1978) Fine (1999) , Williamson (1994, 1998a, 1998b), Williamson and Sainsbury (1998), Sainsbury (1999), Wright (1999), Parsons (2000), Sorensen (1998), Cargile

(1999), and Tye (1999). The view put forth in the present paper is not even mentioned, let alone considered, in any of these works, or for that matter in any of the works known to me on vagueness or the Sorites paradox.

For a review of all the dominant solutions to the Sorites, see Williamson and Sainsbury (1998). The solution I will present is not among those considered by Williamson and Sainsbury.

[308] Obviously I won't be able, in the next few pages, to plug up all the dialectical holes in my argument. A skeptic will thus have plenty of grounds for rejecting what I will say here. So the standard I'd like to be held up to is *not* 'did he prove his case beyond a shadow of a doubt?' but rather 'has he put forth a line of thought that might be worth trying to develop into a cogent analysis?'

[309] The asterices are meant to be what are sometimes called 'quasi-quotes' or 'Quine's corners.' By putting quasi-quotation marks around a piece of verbiage, one quotes everything *except* the variables in that verbiage. In this case, of course, the variable is 'n'. Obviously the verbiage 'an object that is moving n mph is moving fast' is not a sentence proper, because it contains a free variable. So quasi-quotation marks, not quotation marks must be used: by putting quasi-corners around the verbiage 'an object moving at n mph is moving fast', one in effect forms the quotation of any sentence that results from replacing the 'n' with a constant. I couldn't use *actual* quasi-quotation marks because my software doesn't generate them.

[310] From now on, the qualification 'for some choice of "n"' will be left implicit.

[311] So far as I know.

[312] See Fodor (1975), Pylyshin (1984).

[313] It might be appropriate to provide a more thorough statement of the alleged connection between SCT and the computational theory of mind. Computers are information-processing engines. In computers, information is embodied in strings of "code": computers access information entirely through such strings. A string of code is, at bottom, something that stands for something else. Of course, a symbol is, at bottom, something that stands for something else. So strings of code are symbols. So computers access information (representational content) through strings of code and, therefore, through symbols – they think in symbols. So if we are computers, we think in symbols

I myself do not think that computers are, in any genuine sense, information processing engines. I think that is a tendentious and false way of looking at them. I am simply delineating what is taken to be the connection between the computational theory of mind and the symbolic conception of thought.

(I think that "computers" are just physical objects which are conveniently *interpreted* as performing computations and other information-processing operations, but that "computers" no more perform such operations than bananas. Computers do not compute, in my opinion. This, of course, is not a new point of view. Wittgenstein once asked: "Do calculators *calculate*?", the implied answer being "no". But, of course, a proper defense of this position lies outside the scope of the present paper.)

[314] See Fodor (1975: 65-67), Lycan (1984: 237), Kuczynski (2002).

[315] See Fodor's point on page 20 of the present essay; see also Lycan (1984: 247). Kuczynski (2002) argues that the Fodor/Lycan point – that we *use* our internal code without *understanding* it – is inconsistent with the productivity and systematicity of thought.

[316] Wittgenstein (1974: 5): "How can one talk of 'understanding' and 'not understanding' a sentence? Surely it's not a sentence until its understood." Rush Rhees mis-translates "satz" as

proposition. In the context in question, Wittgenstein clearly means “sentence” by “satz”, which is, in any case, what “satz” literally means.

[317] See Fodor (1975: 65-67) and Lycan (1984: 237). See also the quotation from Fodor (taken from private correspondence) on page 21 of the present paper.

[318] This point is made in Searle (1992). Horst (1996) gives an exceptionally thorough and eloquent defense of it.

[319] In *The Language of Thought* and in various other writings, Fodor in effect advocates (S); he says that anything that mediates representations is *ipso facto* a language. (See Fodor 1975, chapters 1 and 2, especially page 55.) This implies that anything that represents anything else is *ipso facto* a bit of language. But this means that sense-perceptions are linguistic *merely* by virtue of being representational. And this, as I’m now trying to show, trivializes the idea that information is always accessed symbolically.

[320] I say *sufficient* condition. I do not think that being intrinsically representational is a *necessary* condition for being mental. In any case, it is a matter of great controversy whether being intrinsically representational is *necessary* for being mental. But it seems to be me quite uncontroversial that being intrinsically representational – as opposed to derivatively representation – is *sufficient* for being mental.

[321] OA stands for “only access.”

[322] It might seem that if sense-datum theories of sense-perception are correct, then sense-perceptions *are* symbolic apprehensions of things. But this is not the case. For the sake of argument, let us suppose that sense-datum theories are correct. (Incidentally, I am open on this question.) Sense-datum theories say that, in having a sense-perception, one is grasping some mental entity – some kind of analogue or icon of an external object. Now this analogue is indeed a symbol. But the sense-perception *of* that analogue is not a symbol. The sense-perception of that analogue is not a grasp of the analogue *by way of* a grasp of something other than that analogue. It is just a grasp of the analogue. So a sense-perception is a completely unmediated and direct grasp of *something* – namely, a certain analogue. Thus a sense-perception is a non-symbolic grasp of *something.*

[323] To avoid verbosity, I will not run a version of this argument for each of these possible substitutions – I will leave it to the reader to make these substitutions and verify on his own that the argument goes through for at least some of them. And we’ve seen that there are non-symbolic representations. So for at least *one* of these substitutions the argument must go through.

[324] Wittgenstein (1953) famously argued that there could be no “private language.” It isn’t clear what he meant. But he surely could *not* have meant that mental images couldn’t serve a symbolic function. For nothing could be more obvious than that people expedite their own reasoning by using mental imagery in a symbolic capacity. So Wittgenstein’s target could not have been that *all* private symbolism is an impossibility; he must have had a narrower target. Perhaps what he meant is that all *verbal* symbolism is necessarily public.

[325] Lycan (1984: 237). Lycan’s emphasis.

[326] In any case, it is hard to see why so much cognitive activity would be involved in the one case and not in the other.

[327] Davidson 1984.

[328] See Davidson 1984, Evnine.

[329] Russell (1903) said that “P is a proposition” is equivalent to “P entails P”. He is right. But that probably doesn’t qualify as an *analysis* of the concept of a proposition. (I don’t think Russell was trying to give an analysis.)

[330] At least among those who concede that there are *any* analytic sentences.

[331] Kripke 1972.

[332] I say a ‘Kripke-style’ argument because the argument I’m about to give *uses* a lot of ideas due to Kripke, and is *ultimately* due to Kripke, but I wish to leave it open whether Kripke himself gave exactly that argument; I want to avoid exegetical controversies.

[333] Due, I believe, to David Kaplan.

[334] One might make the following, hyper-technical objection:

> One does NOT know, wholly on the basis of linguistic knowledge, that “Bush is named ‘Bush’” is true. For if “Bush” doesn’t refer, then that sentence doesn’t encode any proposition, and therefore is neither true nor false (and thus not true). And to know that “Bush” refers, one must have non-semantic knowledge.

There are couple of things to say in response. First of all, to know that “Bush” refers to Bush *is* semantic knowledge. Knowledge of what the referring terms in a language refer to is *quintessentially* semantic knowledge.

Second, even if, for the sake of argument, we implausibly concede that “Bush is named ‘Bush’” is not analytic, that doesn’t matter. Our argument still works against similar sentences that are clearly analytic, like “if Bush exists, then (in English) he is named ‘Bush’”.

[335] What follows borrows heavily from Kaplan 1989. See, in particular, page 509.

[336] Remarks similar to those I’m making here scattered throughout Kaplan 1989.

[337] Kaplan 1989.

[338] There are delineations of the term “language” on which a single language can undergo semantic changes. Since 1700, the English language has gained new words and lost old ones.

[339] For a related point, see Kaplan 1989: 509.

[340] Kripke (1972: 54-56).

[341] Kripke (1972: 56).

[342] Bonjour (1998: 13).

[343] A great deal of the confusion regarding Kripke’s discovery is due to his own choice of nomenclature. When talking about (*), Kripke always uses the term ‘statement’. He doesn’t use the word ‘sentence’ and he doesn’t use the word ‘proposition’. Now the term ‘statement’ is indeed

ambiguous between a bit of verbiage (or a use thereof) and the thing *meant* by that verbiage. I think that Kripke himself was misled by his own word-choice.

[344] Russell held it. He acknowledged that, in some contexts, replacing "Cicero" with "Tully" led from truth to falsity (or vice versa). But he thought (to some extent on independent grounds) that these are quantifiers, not referring terms. And extensionality obviously doesn't demand that truth be preserved when *non*-referring are replaced with non-referring terms.

[345] I say "supposedly" since it is disputed whether definite descriptions are, indeed, referring terms.

[346] I should qualify this statement.

First of all, I grant that *if* sentences referred to truth-values, *then* not all contexts would be extensional. In many contexts, replacing a sentence with another sentence of the same truth-value does *not* preserve truth-value.

But it is not hard to show that sentences do not refer to truth-values. For reasons I give elsewhere, there can be no doubt that sentences *do not* refer to truth-values. The argument used to establish that thesis -- Frege-Church slingshot -- is completely broken. (Barwise and Perry put their finger on the basic problem with it.) Exact analogues of that argument "establish" that there is only one fact, one proposition, one truth, one sentence-connective…And the idea that sentences refer to truth-values butchers any grip we have concerning the nature of reference. But these are points I cannot discharge here. So the present paper shows that extensionality holds, modulo the just made point about sentence-reference.

[347] Of course, there may be other reasons why it would be irrational to do that. I may learn that there is only one great racquetball player in my part of the country. In that case, it would be irrational of me to reject the one while accepting the other.

[348] I think that *all* assertions are statements about propositions. Suppose I say "Plato snored". On the face of it, this is a statement about a person, not a proposition. But in reality it is about both. It is *directly* about a proposition that *concerns* a person. Saying "Plato snored" is the same thing as saying "*that Plato snored* is true". I don't mean merely that these are logically equivalent statements. I mean that they are *the very same statement*; the only differences are phonological. Consider the expression "that Plato snored". Uncontroversially, this denotes a proposition: *that Plato snored.* But "that Plato snored" is not true or false. This is because it doesn't have assertoric force. ("Did Plato snore?" bears the same proposition as "Plato snored". But the former isn't true or false; it lacks assertoric force.) In English, we give assertoric force to the expression "that Plato snored" by dropping the "that" – just as we give interrogative force through certain inversions of word-order. By dropping the "that", I am *affirming*, rather than merely referring to, the proposition denoted by "that Plato snored". So by dropping the "that" --- that is, by saying simply "Plato snored" – I am saying of a proposition (*that Plato snored*) that it is true.

[349] This needs to be qualified. If Russell's theory is right, then definite descriptions do not refer to *individuals.* It could be maintained that, if that theory is right, they refer to 2^{nd} level functions or properties. Here, when I say that "according to Strawson (Russell), definite descriptions (do not) refer", I mean "(do not) refer to *individuals*".

[350] My own view is that they *are* referring terms. First of all, we know that tokens of indexicals and proper names (I mean what *we* call "proper names": "Socrates", "George Bush") are referring terms. Given some results in contemporary semantics, these things are *very* hard to deny. Russell had several arguments that purport to show that definite descriptions are quantifiers. They are good arguments. But exact isomorphs of those arguments "show" that names and tokens of indexicals are quantifiers. They are not. So Russell's arguments must involve a non-sequitur.

The non-sequitur they involve is *basically* the supposition that communicated meaning coincides with literal meaning. There can be absolutely no doubt that "the inventor of bifocals snored" has a very different cognitive value from "the first post-master general snored". And this was (basically) Russell's reason for treating definite descriptions as quantifiers. But an analogue of that arguments shows that "Cicero" and "Tully" are quantifiers. We know they are not. And we can account for the cognitive difference between "Cicero snored" and "Tully snored" while holding onto the idea that they are proper names, not quantifiers. (I develop this idea in___)

Further, as I argue in (___) Russell's theory cannot account for the behavior of definite descriptions in questions and imperatives. It turns questions like "did the inventor of bifocals snore?" into something that is either nonsense or is totally different from what is actually meant by that question.

[351] Or tokens thereof, to be precise.

[352] Barwise and Perry (1983) think that definite descriptions are semantically ambiguous: they are *both* quantifiers and referring terms. Kripke (1977) provides a compelling argument against that view. If "the dog", in the singular, were thus ambiguous, then we'd expect other languages to have *two* words for it (just as other languages have two words for "bank"): one for "the dog" the quantifier, the other for "the dog" the referring term. But it appears that there are no languages meeting this condition.

Keith Donnellan holds that it "the phi" is "pragmatically ambiguous". The idea seems to be this. The word "you" has a uniform meaning; but, depending on the context of utterance, different tokens of it refer to different things. So "you" has a kind of pragmatic ambiguity, even though, strictly speaking, the expression itself (the type, not tokens thereof) has a uniform meaning (some kind of function from contexts of utterance to objects). In Donnellan's view, "the phi" is similar. This, it seems to me, is the same as Strawson's view.

[353] These days, Russellians say that "…the phi…" means: there is exactly one *contextually salient* object x that has phi and…x…

If I say "the dog is sick", I am obviously not saying: there is exactly one dog x and x is sick. Russellians don't want "the dog is sick" to mean: *there is exactly one dog x in all of existence and x is sick.*

So Russellians typically say that "the dog is sick" means: there is exactly one *contextually salient* dog x and x is sick.

A similar strategy employed by contemporary Russellians is to say that "there is" is context-sensitive – it ranges only over a restricted sphere of items. (If I say "there's no beer" I don't mean there is no beer in all of existence, but only in our house.)

So according to this view "the dog is sick" means: there is [within a certain restricted context] exactly one dog x, and x is sick.

Referentialists make a corresponding adjustment to their view. The referentialist says that "the dog is sick" doesn't presuppose the existence of exactly one dog *in all of existence*; it presupposes that there is exactly one *contextually* salient dog.

[354] Of course, in some social contexts, definite descriptions come to function as nick-names and may, in other respects, be used in non-standard ways. So and so's nickname might "the terminator" or "the master". Obviously, in that kind of context, "…the terminator…" does not bear the proposition : *exactly one x is a (contextually salient) terminator and…x…*But, there, obviously we are either dealing with definite descriptions that have "grown capital letters" and have become proper names *or* we are dealing with non-literal uses of definite descriptions.

Similar considerations deal with Donnellan's case. If I say "the man drinking the martini snores", and nobody (contextually salient) is drinking a martini, then surely what I have said is, at the level of literal meaning, not true. Of course, people may know well enough what I *meant* – namely *so and so snores*. But it would be very hard for *anyone* – whether Russellian or referentialist – to make a case

that, under that circumstance what I said was literally true.

[355] In any case, it won't be true absent further information about *which* of those eighty people is meant. *Some* restriction of context – *some* making-salient of one of those lunch-eaters – must be effected. We will deal with issues concerning contextual salience in due course.

If I say "the man who ate lunch today in Asia snores", you need to know which *particular* lunch-eater in Asia I have in mind. Some *particular* such individual needs to be made salient. If I say "I don't have any *particular* such man in mind" – if I don't make *one* such lunch-eater be salient – then my utterance is surely not *true*.

[356] Obviously Kaplan's two great articles – "Quantifying In" and "Opacity" – deserve to be mentioned here. In the former, Kaplan is operating within a Fregean framework (the semantic values of names are taken to Fregean senses, and *not* individuals). This fact, quite frankly, nullifies the parts of his analysis that are relevant in the present context. In "Opacity", Kaplan is operating within a direct-reference framework. But he still doesn't quite provide an intuitively satisfying of the difference in cognitive value between (**) and (***).

[357] Of course, the term "propositional guise" is *not* dubious if it refers to the existential information through which, according to my theory, one grasps singular propositions.

[358] There are two points I think worth making, but not important enough to include in the main text.

The first point has to do with the relation of my views on indirect discourse to a view that has a certain prevalence in the existing literature.

The second point has to do with the fact that Russell's theory is seldom held in quite the form in which I've presented it: some subtleties relating to *contextual salience* are an integral part of contemporary Russellian views on definite descriptions. But it is easy to extend what I said about definite descriptions to accommodate this fact.

Point #1

A metalinguistic account of apparent substitution failures in indirect discourse

I hold, with many others, that a name's semantic content is its referent. I have argued that, if "Smith" and "Bob" are co-referential, then

(*) "Jones believes that Bob is a great racquetball player"

has a true proposition for its literal meaning iff

(**) "Jones believes that Smith is a great racquetball player"

has a true proposition for *its* literal meaning.

This position is widely held.

It is a commonplace that a rational person might assent to (*) without assenting to (**) or *vice versa.*

I would like to discuss what I think is a *false* analysis of this fact – what is a *false* analysis of the cognitive difference (the difference in respect of what is communicated) between (*) and (**).

That analysis is given by the following passage (Moore 1995: 114-115):

Recently the view that the semantic value of a proper name is always just its referent has come to be fairly widely held. Barwise and Perry (1983), Kaplan (1977), and Soames (1987) have all been among

the advocates of this position, which has come to be called the "direct-reference theory". According to these theorists, "John believes that Cicero was a Roman orator" actually does entail "John believes that Tully was a Roman orator", but there is a pragmatic reluctance to substitute "Tully" for "Cicero" because we tend to make belief-reports do double-duty – not only reporting what propositions are believed, but also what sentences are accepted. Thus, "John believes that Cicero was a Roman orator" not only reports that John believes that the proposition that Cicero/Tully was a Roman orator, but it also leads us to expect that John would accept as true th sentence "Cicero was a Roman orator." The reason we are reluctant to substitute "Tully" for "Cicero" is that it would be misleading, not that it would be
[358]
false.

Of course, I agree that "John believes Cicero was wise" semantically encodes the same proposition as "John believes that Tully was wise". But with that qualification, the view just described seems to me quite false. In any case, it definitely doesn't apply to all, or even most, cases of apparent substitution-failures in epistemic contexts. Again, consider the sentences:

(t) "Fred believes that Bob is a great racquetball player"

(u) "Fred believes that Smith is a great racquetball player".

Under the circumstances described above, (t) will communicate either:

(w) *Fred believes that: a moment ago, I just shook hands with some person x with such and such properties and x is a great Racquetball player*

or

(x) *a moment ago, I just shook hands with some person x with such and such properties and Fred believes that: x is a great Racquetball player*

By contrast,

(u) "Fred believes that Smith is a great racquetball player".

will communicate either:

(y) *Fred believes that: yesterday, from a great distance with thus and such properties, I saw somebody x and x is a great racquetball player*

or

(z) *Yesterday, from a great distance, I saw somebody x with thus and such properties and Fred believes that: x is a great racquetball player.*

Obviously the first pair of proposition is completely different form the second pair. That is why a person (who is rational and is a competent speaker of English) might assent to (t) and not assent to (u), or *vice versa.*

Notice that none of (w), (x), (y) or (z) have anything at all to do with Fred's beliefs concerning *expressions.* None of those four propositions is meta-linguistic to any degree.

Mutatis mutandis what we said about (t) and (u) applies in spades to

(#) “Fred believes that Superman is strong”

and

(##) “Fred believes that Clark Kent is strong”,

and also to

(^) “Fred believes that Hesperus is lovely”

and

(^^) “Fred believes that Phosphorous is lovely”.

If I am willing to utter (#) but not willing to utter (##), that does not, at least not typically, have to do with my beliefs as to which *sentences* Fred assents to.

Suppose I know Fred to be an Albanian, who speaks not a word of English. But, as I know, Fred is an avid astronomer: he knows the Albanian words for the stars, planets, and so forth. *But* Fred has not had the good fortune to learn that famous astronomical discovery that Frege, and so many other philosophers, discuss.

Under those circumstances, I could very well be communicating something true if I said

(&) “Fred believes that Hesperus is lovely but Fred does *not* believe that Phosphorous is lovely”.

(Again, I grant that the proposition semantically encoded in (&) is self-contradictory *at the level of literal meaning.* But what it communicates need not, and probably is not, self-contradictory.)

Now what I am communicating by saying (&) is not about Fred’s attitudes towards *expressions.* Fred has never even heard the expressions “Hesperus” and “Phosphorous”. What I am communicating is about Fred’s astronomical beliefs. I am *communicating* that Fred has certain attitudes towards certain existential, but not meta-linguistic, propositions.

Of course, one might say: “Maybe what you are communicating is Fred’s expression towards certain *Albanian* expressions”. But this seems to me quite false; we are talking about Fred’s astronomical beliefs – not his beliefs about nomenclature, not even nomenclature relating to astronomy. When we say things like (to switch examples momentarily)

(@) “the Babylonians believed that Hesperus was lovely, but didn’t think the same of Phosphorous”,

we are surely not talking about the Babylonians’ attitudes towards *expressions*. Of course, they (and Fred) *have* expressions corresponding to our terms “Hesperus” and “Phosphorous”. But the discovery that we report by saying “Hesperus is Phosphorus” is not a discovery about expressions. When we say that the Babylonians *failed* to make this discovery, or that Fred *failed* to make it, we are not talking about their failure to make some *verbal* discovery. When we say things like (&) and (@) we are not talking about their attitudes towards *expressions*; we are talking about decidedly *non*-metalinguistic beliefs. We are talking, if my analysis above is correct, about their relations to existential, but not typically metalinguistic, propositions.

Suppose that Fido is your dog. Sometimes (for some odd psychological reason that we don’t have to go into here) you dress up in a scary costume, including a creepy mask, when you play with Fido. When you engage in these antics, you address yourself as “Argo”. (So, to taunt poor Fido, you say “I,

Argo, am going to get you! Hardy har har!") Other times you play with him in your regular clothes and don't wear a mask. Fido, let us suppose, doesn't know that there is some *one* individual who wears a scary mask while playing with him and also sometimes *doesn't* wear that scary mask while playing with him – Fido is taken in by your ruse. You would, no doubt, be communicating a perfectly correct proposition if you said (to someone who was familiar with your strange antics):

($) "It is not the case that Fido likes Argo, but Fido does like Larry".

(You are Larry, for the purposes of this story.) What you are communicating by saying ($) (not what is semantically encoded in it) is surely *not* some proposition about Fido's attitude towards *expressions.* Fido may not have any attitudes towards any expressions. What is communicated by ($) might be true even if Fido had never heard a word in his life. The difference, in respect of what is communicated, between

($$) "Fido does like Larry"

and

($$$) ""It is not the case that Fido likes Argo"

doesn't have to do with Fido's meta-linguistic proclivities or with your *beliefs* as to his (perhaps non-existent) metalinguistic proclivities.

The difference *does* have to do with Fido's assenting to the proposition

(!) *there is this weird guy who wears a mask and plays with me, and I don't like him*

and his *not* assenting to the proposition

(!!) *there is this nice guy who gives me food and plays Frisbee with me, and I like him.*

I'm not ruling out the possibility that in *some* cases the metalinguistic analysis given by Moore is the right one. But it is decidedly implausible to suppose that his analysis corresponds to what is typically, or even more than very occasionally, going on. We have provided a way of dealing with the phenomenon in question – of dealing with *apparent* substitution failures in epistemic contexts – that, so far as I can tell, is consistent with direct-reference theory and also with certain key epistemological facts. So far as I can tell, there is not much room in that account for the kind of metalinguistic story that Moore is giving us.

In the *Begriffsschrift*, Frege says that "A=B" is informative because it tells one that "A" and "B" co-refer. Later, in "On Sense and Nominatum", Frege rightly rejects that theory. The discovery that we (infelicitously) express with the words "Hesperus is Phosphorous" is not linguistic at all. The theory that Moore describes is, it seems to me, a variant of the *Begriffsschrift* theory, and must be rejected for the reasons that theory is to be rejected.

Point #2

contextual salience and neo-Russellian theories of descriptions

It certainly *seems* that "John thinks the phi is ill" may differ in truth-value from "John thinks that the psi is ill", even if the phi *is* the psi. We gave an argument earlier showing why, *even if definite descriptions are referring terms,* there is no reason to think that this apparent substitution failure is an *actual* one – there is no reason to think that those two sentences encode *different propositions.*

But the argument we gave is really a first approximation. To make it more cogent, we must register a certain point about Russell's theory. Suppose you and your wife have a dog. One day you find your dog ill. You say to your wife:

(ix) "The dog is sick."

If we go by what Russell himself said, then (ix) means: *there is exactly one dog x and x is sick.*

But it would be hard to maintain that, when you said (ix), you were asserting that there was exactly one dog in all of existence. (That said, this is actually what Kent Bach (1987) maintains.) [358]

What contemporary Russellians (see MacCawley 1983, Neale 1990) [358] usually say is this: (ix) means:

There is exactly one contextually salient dog x, and x is sick

or

There is in some restricted context *exactly one dog x, and x is sick.*

This seems like a very reasonable adjustment to Russell's theory. If this is right, then Russell's theory is really this.

"…the phi…"

means:

there is exactly one contextually salient x, and…x…

So

(ix) "the dog is sick"

means

[358]

(x) *there is exactly one contextually salient dog x and x is sick.*

And the referentialist has to make an analogous adjustment to *his* position. The referentialist will not, at least not typically, say that "the dog is lovely" is abortive if there is more than one dog *in all of existence.* At the same time, everybody – referentialist and Russellians alike – think that "the" somehow connotes uniqueness. (It is not like "some" or "three".) The referentialist says that, if "the dog" is to be non-abortive, then there must a dog that is, from some viewpoint, unique. Of course there are many dogs, and tokens of "the dog" are not typically abortive. So the referentialist must presumably mean that, if a token of "the dog" is to non-abortive, then there must be a unique *in some particular*

context – the one relating to the context of utterance . In any case, this is the natural way to interpret the kind of uniqueness that is, from the referentialist's viewpoint, involved in the word "the".

So here is what the referentialist says about the semantics of "the dog". It is a referring term. If a token of "the dog" is to be non-abortive, there must be at least one contextually salient dog and at most one such dog. Supposing that O is the contextually salient dog, then "…the dog…" bears the position: …*O*…So under those circumstances, a token of "the dog looks ill" bears the proposition: *O looks ill.*

So for the referentialist, the semantic rule for "the dog" is:

If there is exactly one salient dog x in some context C, then a token of "…the dog…" in C bears the proposition:…*x*…

Of course, if the referentialist is right, then supposing that people have some (minimal) awareness of the semantic rules constituting their own language, then any one hearing an utterance of (say) "the dog is sick" will know that that utterance is true iff

(1) there is exactly one contextually salient dog x, and x is sick.

So, if the referentialist is right, then a token of "the dog is lovely" will *communicate* (1), even though it won't semantically encode that proposition.

So the referentialist does *not* have trouble accounting for the fact that "the dog is sick" communicates (1) – that "the dog" *acts* like a Russellian quantifier. And, for reasons we've seen, this fact reconciles the assumption that "the dog" is a referring term with the *appearance* of a possible difference in truth-value between "John thinks that the dog is sick" and "John thinks that Fred's favorite animal in the whole world is sick".

[359] I believe that Wittgenstein denies it in his *Remarks on the Foundations of Psychology*. With all due respect to that great author, his grounds for this *prima facie* highly counter-intuitive position are, at least to me, very obscure.

[360] But as Kripke (1972) and Kaplan (1989) have observed, the converse is not always true. Kripke and Kaplan have pointed out (or at least compellingly argued) that some sentences are analytic *and* express a posteriori propositions. So while all sentences that express a priori propositions are analytic, the converse is not true. In another paper, I say how analyticity *is* to be defined, given that it isn't quite identical with the property of bearing an a priori proposition.

[361] Boghossian (1997: 331).

[362] Dummett 1978.

[363] Quine's nominalism is not a peculiarity of *Quine's* version of extreme empiricism. Nominalism is, at least arguably, a necessary commitment of any extreme form of empiricism. A non-nominalist is *ipso facto* committed to the existence of non-spatio-temporal entities. Such entities are presumably not known by studying the spatio-temporal world. (If you try to learn about a platonic entity by looking at a lot of non-platonic entities, you're looking for answers in the wrong place.) It follows trivially that abstract (non-spatiotemporal) entities must be known through (to paraphrase Descartes, who speaking in a different context) a purely intellectual scrutiny. So Quine's dismissal of propositions, and his willingness to talk *only* about sentences, and not about sentence-*meanings* (propositions, is a fairly direct consequence of his empiricism. Thus in showing the absurdity of Quine's nominalism, we are not producing some purely *ad hominem* argument against Quine, but an argument that applies to empiricism generally.

[364] There is another counter-attack. Obviously not *all* sentences are analytically equivalent. (Obviously Quine will agree with that since he doesn't believe in analyticity at all, and therefore doesn't believe in analytic equivalence. So we are not begging any questions against Quine here.) So there is at least one pair of sentences (S1, S2) that are not analytically equivalent. Now something has to *follow* from the one sentence which doesn't follow from the other. Otherwise, they *will* be analytically equivalent. (After all, by definition, two sentences are analytically equivalent iff nothing follows from the one which doesn't follow from the other.) But if something follows from the one which doesn't follow from the other, then of course something must *follow* from at least one of them. In which case *there is* analytic truth.

Actually, these considerations suggest a way of proving that there are some analytic truths. Quine can't concede that *any* two sentences are analytically equivalent. He must hold that, for any sentences S1 and S2, it is *not* the case that something follows from the one which doesn't follow from the other. So Quine must hold that, for any two sentences, *something* must follow from the one which doesn't follow from the other. But in that case of course, he is committed to analyticity.

But, I admit, this argument doesn't have too much force. For Quine could say: 'All right. I believe that *nothing* follows from any sentence. So *technically* I am committed to holding that all sentences are, vacuously, analytically equivalent. But in being thus committed, I am somehow not *substantively* committed to analytic truth.' I do think this is a viable rejoinder for Quine; hence I have put this line of attack in a footnote.

[365] See McCawley (1981:15-16) for a compelling argument to this effect.

[366] I mean, he denies that there are concepts in the platonic, not necessarily the psychological, sense.

[367] Bonjour (1998: 76-77; 90-96) gives exactly this argument.

[368] Dummett 1973: 592-596, especially the first full paragraph of 596; also Dummett 1978.

[369] The argument I'm about to give is very similar to one given in Dummett: 1973: 592-596. What I say here is also similar to an argument given by Bonjour 1998: 5; 76-77; 90-96.

[370] The important point that propositions affirming probability-relations is made in Bonjour (1998: 206-217). It forms an integral part of his solution to Hume's problem of induction: a solution which, I think, goes through. Dummett (1978: 377, 380-381) makes essentially this point.

[371] Bonjour (1998: 91) says something similar.

[372] Bonjour (1998:5).

[373] This does not mean that whenever someone believes something on the basis of memory, he is ipso facto making an inference. (When a three year old believes that e.g. his mother has red hair on the basis of memory, he is probably not *reasoning* to a conclusion from some premise concerning the reliability of his memory.) It means that, if he were to give the *justification* for any such belief, he would (if he were rigorous about it) have to acknowledge the existence of an inferential component like the one described.

[374] Pap made a very similar point in response to Leibniz's contention that all a priori truth "reduces" the "law of identity" (everything is self-identical). Pap points out that one *needs* a priori truths other than that law to reduce (say) *there are more reals than rationals* to an instance of that law. (Pap 1958: 11-12).

[375] Fodor (1998: 109) makes this point. I think it ultimately goes back to Wittgenstein (1974).

[376] I am using the word ‘concept’ in its non-psychological sense. So I’m using it in the sense that it bears in the sentence ‘for any x, if x fall under the *concept* TRIANGLE, then x falls under *concept* CLOSED FIGURE.’

[377] Quine (1968: 59).

[378] See Brandom (1998: 88-89).

[379] Of course, this point was famously made by Wittgenstein (1953).

[380] I must point out that Davidson himself accepts – or *claims* to accept – that Quine *has* demolished the analytic-synthetic distinction. I think that Davidson is being inconsistent with himself since, as we’ll now see, a lot of his work implicitly relies on that distinction.

[381] Davidson (1994: 232).

[382] Though it is arguable and argued against.

[383] Crispin Wright gives an argument similar to the one just given. (I wrote the above before I encountered Wright’s argument.) See Hale (1998: 492-493).

[384] Of course, Carl. Hempel (1945a, 1945b) discovered the paradox. The paradox has been re-articulated many times, sometimes with interesting twists. (See Maher 1999). Hempel bit the bullet, saying that one *does* confirm (R1) by encountering non-black non-ravens. An intuitively more plausible view is proposed by Quine (1969). (My solution here has some vague affinities with Quine’s view.) Some sharp -- in fact, devastating -- criticisms of the views prevailing during his time, including Quine’s, is given by Ayer (1971). Scheffler (1981) admirably reviews all the major solutions prevailing at the time of his writing. Most solutions propose that (R1) and (R2) are, indeed, equivalent. See Schlesinger (1974), Trusted (1979), Cohen (1989), Leavitt (1996), Sainsbury (1997), Maher (1999).

There are views that try to destroy the alleged equivalence between (R1) and (R2). But these views always resort to non-classical logics. (See Sylvan and Nola (1991)).So in effect these views *do* concede that, within a classical framework, (R1) and (R2) are indeed, equivalent. So these views presuppose that (R1) and (R2) *are* equivalent, if only within a classical framework: otherwise non-standard logics wouldn’t be invoked to block the equivalence.

These are all worthy views. For reasons of space I cannot review them all here. But what I can say is that, since they all presuppose a certain equivalence between (R1) and (R2), they are hobbled at the outset. Again, I grant that on *a* legitimate reading, (R1) and (R2) are, within the framework of classical logic, logically equivalent. But “all” statements are not always so transparent in terms of their content. And this is a case where the nuances of the word “all” are relevant. I submit that once we make these nuances explicit, we will find that *even within a classical framework*, what is said (or at least meant or insinuated) by (R1) and (R2) is not equivalent.

[385] Hempel (1966).

[386] Of course, Hume denied that there are nomic relations. (Or, what is the same, he said that nomic relations just *are* chance concomitances.) But Hume also said that the past is no basis for prediction. Hume was being consistent. *If* Hume’s analysis of natural law is correct – if laws are *de facto* constant conjunctions – then every new event is a fresh roll of the dice; and one would be guilty of a version of the so-called gambler’s fallacy if one were to make predictions about future dice-rolls on the basis of past ones.

[387] The quantifier is subject to implicit restrictions. There may be Martians that can drink arsenic without dying.

[388] It might be said that the connection between something's being water and its being composed of oxygen is not *nomic* – that it is better described some other way. Perhaps the term "nomic" should be reserved for contingent relations.

In my view, such a position would prejudge many important questions about the relation between nomic and metaphysical necessity – about the grounds of nomic necessity itself. A number of propositions that we intuitively regard as nomic are not contingent. If we removed all non-contingent propositions from scientific theory, it would be very difficult to organize what was left over. Non-contingent propositions, such as (WO), are such an integral part of the theories that describe natural law that it would be artificial to deny that they had nomic status.

Kripke (1972) describes the relation expressed by (WO) as "metaphysical". There is no metaphysically possible world where water doesn't consist (in part) of oxygen. I agree with Kripke. But it doesn't necessarily follow that metaphysically necessary relations are not also nomic. As a matter of simple fact, they often serve an explanatory function similar to that of principles that are *plainly* nomic. In the third lecture of *Naming and Necessity*, Kripke himself suggests that, in some cases, nomic relations might be *necessary* in the strictest sense. So I don't think I am completely off-base in describing the relation between the property of being water, on the one hand, and the property of being composed of oxygen, on the other, as "nomic".

[389] This doesn't mean that instances of phi cause there to be instances of psi (or vice versa). It could there, for some third property chi, it is instances chi that produces both instances of phi and of psi. But, in this case, the relationship between instances of phi and of psi, though not itself one of causation, would nonetheless be causal: the concomitance of phi's and psi's would have to be explained causally (though not as a causal relationship between phi's and psi's).

[390] For the record, I am fairly certain that the *causal* story is the right one. The relation between being a raven and being black is not like the relation between being water and being composed of oxygen atoms. The first relation, unlike the second, seems to be contingent, even though it is a principled one. We could genetically engineer white or green ravens. This would involve tinkering with the causal mechanisms mediating between the properties distinctive of ravens, on the one hand, and their phenotypic expression, on the other.

[391] short for "Claim #1."

[392] Short for "celestrial object."

[393] Salmon (2007).

[394] As McCready notes, Chomsky (1957) used this very sentence. As I recall, however, Chomsky uses it to show that sentences that hitherto haven't been used *do* have meanings. But if McCready is right to impute this view to Chomsky, then Chomsky is wrong, and for quite a variety of reasons, as we'll see. (Which is why I haven't bothered to see if, in fact, McCready's Chomsky-exegesis is correct.)

McCready's wording is ambiguous; and no sooner is it disambiguated than the problems with McCready's contention begin to emerge. McCready says that CG has no "coherent meaning," which, he claims, decisively shows that *bona fide* sentences can be meaningless.

When McCready says that CG has no "coherent meaning," what is he saying? If he's saying that it *does* have a meaning, then his point defeats itself. So that's not what he's saying. He must be saying

that, although its various parts have meanings, those various meanings fail to cohere into a single proposition (or, in any case, larger meaning of some kind or other). In other words, unless McCready's contention is very obviously false---unless, it's only a slight exaggeration to say, it's to the effect it is itself false----it must be to the effect that *nothing* is affirmed by (tokens of) CG. Which is what I will take McCready to be contending, both because, exegetically speaking, it clearly is what he's saying and, second, because a scholar of McCready's stature wouldn't make a brazenly self-defeating claim.

[395] Also, quite literally nothing that I say in my book depends whether the claim McCready is attacking is true. McCready is attacking a complete straw man. (See my forthcoming counter-review, in this journal.) I do make the totally innocuous claim that, if produced by a volcano (50 million years ago) or a coffee-machine (today), a noise that was acoustically indistinguishable with "snow is white" wouldn't be a *bona fide* sentence-token. But that's as close as I come to saying anything that coincides with, or depends on, the thesis against which McCready is inveighing. That said, McCready's claim is widely accepted, as are his reasons for accepting it. So this claim, and these reasons, are worth examining.

[396] My inclination is to think that, *at the level of semantics*, these sentences don't have logic or nomic force, and it is through implicatures, not explicature, that utterances of them transmit the logically/nomically pregnant claims just describes. I say this because those implicatures, if that's what they are, can be cancelled and are therefore, by Grice's (1975) plausible criterion, pragmatic epiphenomena, as opposed to semantic explicatures. Cf.

> "All metal objects expand when heated; but there's nothing inherent in what it is to be a metal object that disposes such objects to expand when heated. Long ago, hoping to mislead future scientists, some alien prankster, hoping to mislead future scientists destroyed all metal things that didn't expand when heated."

Most semanticists and pragmaticists accept Grice's (1975) views on cancellation. Wayne Davis rejects it. But he's an exception. Which doesn't mean he's wrong, of course---only that, by assuming *arguendo* the falsity of Davis' position, I'm arguing within a framework within which a lot of important linguistic and philosophical work is conducted and whose hypothetical correctness it is not unreasonable to assume.)

But, of course, it's a delicate question whether, in some contexts or some values of phi and psi, ‹all phi's and psi's› is, *at the level of literal meaning*, to the effect that a thing's a being phi is connected in some robust, not purely statistical way, to its being a psi.

[397] The distinction between the logically false and the logically ridiculous is a purely psychological distinction. Thanks to Gödel (1931), we know that

(i) "arithmetic can be formalized"

is false. But (i) isn't ridiculous, at least not to us. However,

(ii) "there are square circles"

is ridiculous.

But to beings much more intelligent than ourselves, (i) wouild be quite as ridiculous as (ii), the same being true of:

(iii) “there are regular, 17-sided polygons”

and

(iv) “if the Theory of Types is true, any attempt to affirm it is false, wherefore it’s false.”

If, like McCready, we say that anything that, like (ii), “has no coherent meaning” is *ipso facto* devoid of meaning, then we’re saying that each of (i) and (iii) and (iv) is devoid of meaning; and, by implication, we’re saying that it is actually impossible to say anything that isn’t logically, mathematically, philosophically, or otherwise non-empirically false. Which is itself non-empirically false. And ridiculous.

[398] Russell, B. (1940), p. 222. See also Scheffler, I. (1981), p. 159.

[399] Also, procrastination is a property of certain modes of conduct; and, being a property, it’s non-spatiotemporal and therefore can no more be drunk than quadruplicity can drink. To be sure, *instances* o procrastination are spatiotemporal. (My putting off making my bed is something that happens, or doesn’ happen as it were, in space-time.) But

(QP) “quadruplicity drinks ionstances of procrastination”

is false, since quadruplicity doesn’t drink anything, as previously noted, and instances of procrastination not being liquids, cannot be drunk. So MS and QP are deeply idiotic, and therefore absurd in some sense of the word, but perfectly meaningful.

[400] Ryle is, I suspect, quite right to say that, because it is to be understood in organizational (functional), not (strictly) geographical terms, Oxford University could, without ceasing to be Oxford, relocate itself. There would have to be some sort of causal and functional nexus linking the pre- and pos repositioned entities for them to be constitutive of some one university. But that’s consistent with Ryle’s position and, in any case, isn’t relevant here, since we’ll grant Ryle’s premise and show that, relative to Ryle’s argument still fails.

[401] The brackets are meant to be quasi-quotes.

[402] Scheffler (1981), p. 159.

[403] “BW” is short for “boiling water.”

[404] “GW” is short for “glubbified water.”

[405] “BW” is short for “boiling water.”

[406] “GW” is short for “glubbified water.”

[407] Scheffler says that GW is meaningless *with respect to E*---where E is a hypothetical language such that no sentence S belongs to E if S is non-empirical (or non-tautologous). But Scheffler doesn’t say, or acknowledge the possibility of saying, that GW is meaningless in some less conditional sense.

[408] There are obviously sentence-*types* that have no truth-conditions, e.g. "he laughed at her." In fact, in natural language, few if any sentence-types are entirely context-insensitive, given that few if any of them lack tense-markers.[408] But the logical positivists rightly disregarded the likes of "he ate it" and "she laughed derisively at her butler." And they were right to do so: when considered in a vacuum, "he laughed at her" can be treated as an open-sentence, the "he" and "her" being (restricted) free variables.

We'll see that there are *no* meaningless sentence-types or sentence-tokens. If E is an expression-type such that there is no rule saying what is meant by a given token t of E, then E is without meaning. If there is such a rule, then E is meaningful. If t is an a sentence-token that is assigned *truth-conditions*, even if t fails to have been assigned a proposition, t is still meaningful. If, while pointing at an empty space, I say "that elephant is a genius," my utterance is neither true nor false, since (if Strawson (1950) and Kaplan (1989) are right, as I think they are) the semantic rule for the corresponding sentence-type is:

> (SR) If, in context C, there is exactly one contextually salient elephant, and t is a token in C of "that elephant is a genius," t is true iff O is a genius and false if O is not a genius; if, in C, there is no contextually salient, t is without truth value.

But even though there is no elephant in the vicinity, t still has truth-*conditions*; that is why t---*that very token*---*could* have been true. It would have been true in a world otherwise exactly like our own in which there was a uniquely contextually salient elephant, which, moreover, was a genius. Also, in this world, when trying to determine whether a given token, such as t, is true, false or neither, we start our starting with its truth-conditions---with its *meaning.* (In what follows, "sentence-token" is to mean "sentence-token in the indicative mood.") For a sentence-token to be meaningful is not for there to be some proposition that it affirms (or even expresses); it is the semantic rules of the language to which it belongs to say what *must* be the case *if* that token is to be true. In other words, it is for that token to have truth-conditions. t does have truth-conditions. And, be it known, those conditions do not coincide with the semantic rule for SR. First of all, SR, unlike t, isn't true or false under *any* conditions, not even counterlogical ones. (Whenever I refer to hypothetical conditions, I am assuming that, in them, English is spoken. More generally, any mention, implicit or explicit, of a "possible world" or hypothetical scenario is followed by a (phonetically unrealized) token of "that is semantically just like our world, i.e. that is governed, in the same places, to the same degree, in the same way as our world." For it's obvious, and irrelevant in this context, that a token phonetically just like t could mean anything or nothing.

[409] A technicality: the sentence "some people smoke and some people play squash" is not, technically, a quantified generalization, since the main connective is not "some people" (or any other quantifier), but rather "and." So it's a non-quantified molecular sentence in which two quantified sentences are embedded. In any case, such sentences are covered by what we'll say in Sections 3.2 and 3.3.

[410] According to Russell (1905, 1918) and Frege (1884), the logical form of "some person smokes" is "for some x, x is a person and x snores." The latter sentence can be interpreted as having the form ‹x has phi›. (Replace "x" with a noun denoting the property of being a person who smokes, and replace phi with an expression that denotes the property of being instantiated.) Thus, Frege's insight isn't necessarily represented as being to the effect "someone smokes" does *not* have the form ‹x has phi›; it can be interpreted as saying that it *does* have that form, *provided* that the grammatical subject and

object (“someone” and “smokes,” respectively) switch places (to put it loosely but suggestively) with logical subject and predicate (“the property of snoring” and “is instantiated,” respectively.)

[411] Sometimes the logical positivists described *actual* sentences as meaningless. (Cf. Ayer (1936), Bridgeman (1953), Carnap (1953a, 1953bb).) For example, they claimed that statements about unobservables (e.g. particles, or other peoples thoughts) are meaningless *unless* construed as statements about observables (meter-readings, overt-behaviors). So “Sam is thinking about the real number system” is either meaningless or it’s a statement about his overt behavior. But no statement about Sam’s overt behavior is equivalent with any statement about his overt behavior, since, given any thought T that Sam might be having, and given overt behavior B, there is some conceivable circumstance C such that, given C, along with the fact Sam is having T, Sam’s overt behavior might be B. And no statement about meter-readings (or other observables) is equivalent with one about electrons (or other unobservables), since given any micro-event m, and given practically any observable event E, there is some conceivable instrumental apparatus such that m’s occurrence should lead to E). So in those cases where the logical positivists *did* describe *bona fide* sentences as meaningless, those sentences were meaningful, their dismal attempts to show otherwise veritable proof of that.

[412] See Strawson (1950), Perry (1977), Barwise and Perry (1983), and Kaplan (1989). Strawson made most of the points that, in later years, would constitute the better part of the vigorous debate concerning the semantics of indexical (context-sensitive) sentences, even though, curiously, his name was almost never mentioned.

[413] Short for “that elephant.”

[414] Short for “semantic (rule for that) elephant.”)

[415] In this context, we’re focusing on indicative sentences. But what we’ll say is easily adapted to deal with non-indicative tokens---better adapted, as we’ll see, than the contention than the contention we’re now in the process of refuting, namely that the meanings of sentences are propositions, not truth-conditions, and therefore that, in an elephant free context, a token of TE is without meaning.

[416] Be it noted that, given any sentence-type that contains any context-sensitive component---an indexical (e.g. “now”), a “demonstrative (e.g. “that”) or a tense-marker (e.g. “-ed”)----that sentence-type’s meaning cannot be a proposition. The meaning of “it’s cold” cannot be a proposition, since non-simultaneous tokens of it (that occur in the same place or in different places) may have different truth-values, and must therefore express different propositions (or, at least, have different truth-conditions), and simultaneous tokens of it, if tokened in different places, may have different truth-values. (Given some instance *in*, a token of “it’s cold” that occurs in a sauna, in a dialectical context that makes it clear that it is the sauna’s temperature that is relevant, may be false, even though a simultaneously occurring token that occurs in the frozen tundra outside the sauna (in a dialectical context that makes it clear that it is the tundra’s temperature that is relevant) may be true.

Since every sentence of natural language contains a tense-marker, every such sentence is, at the strictly semantic level, context-sensitive. (In some cases, the context is so blatantly irrelevant that the tense-marker is innocuous (e.g. the tense-marker in “2 divided by itself equals 1”) seems to do nothing. But supposing *per impossibile* that people believed (as some do, and as some have) that mathematical reality changed, the tense-marker in that sentence would not be vestigial. And, were it to become widely accepted that mathematical reality was as changeable as the weather, even sentences like “2/1=1” and “2+2=4”---the formal analogues of “2 divided by itself equals 1” and “2 plus 2 equals 4”---might come to be seen not only as having tense-markers, but as having non-redundant ones. In any

case, setting aside certain (numerically inclusive but) narrowly defined and singular classes of sentences, it's fair to say that most if not all natural-language sentences (sentence-types) contain context-sensitive components and that, for this reason, the meanings of such types are rules of the previously-discussed kind.)

Kaplan (1989) refers to the meaning of a sentence-type as its character; and he refers to the proposition assigned to a given token of a given type as that token's "content." According to Kaplan, given some sentence type T, T's meaning is a rule that assigns *propositions* to tokens of T. I submit that T's meaning is a rule that assigns *truth-conditions* to tokens of T. I further submit that, for a sentence-token to be meaningful is for it to have truth-conditions and is *not* for there to be some proposition that it affirms. (A token's affirming some proposition is sufficient but not necessary for its being meaningful and, what is related, is a consequence, but not the essence, of its being meaningful.)

[417] Notice that (*) doesn't coincide with the meaning of K_1 (the *type* of which k_1 is a token). K_1's meaning is given by a rule that doesn't mention k_1 or any other specific token of K_1.

[418] In any case, this is the Strawson-Kaplan view, and for many reasons—see Kaplan (1989)---it is a reasonable one.

[419] Advocates of CW (e.g. Russell (1908), Tarski (1930), and Kripke (1975)) take it to be to the discredit of the English language that the rules constitutive of it permit the construction of such a sentence. But, those rules being what they are, it is indeed a *bona fide* sentence-token of English. xxx

[420] Not, *nota bene*, the expression that results when it is flanked in quotations; not, in general, any expression *referring to it*, whether homonymous with it or not; but that expression itself. In CW, tf *per se* occurs.

[421] Given *only* what t* means, no *empirical* work need be done to establish that it isn't true. But that doesn't mean that t* is false solely by virtue of what it means. The statement:

> (A) "Establishing that sentence-token s cannot be true involves no *empirical* knowledge (over and above that involved in understanding s to begin with)"

doesn't entail

> (B) "s's failure to be true is strictly a matter of language."

What does entail (B) isn't (A) alone, but rather (A) *conjoined* with:

> (C) "all truths are either spatiotemporal or semantic."

But (C) is false. Quine (1929) and Gödel (1931, 1953) made it clear that, even though it's a matter of semantics' that "3" refers to 3 and that "27" refers to 27, it isn't a matter of semantics that, *given* those semantic rules, "3" refers to the cube-root of the thing referred to by "27." The consistency of a given set of conventions (be they semantic rules or conventions of some other kind) isn't a matter of convention; it's a matter of non-conventional, non-spatiotemporal, logical fact.

[422] *A defense of this claim*: No series of noises is a sentence-*type*: an ink-deposit is, at most, a sentence-token. (In this context, "burst of noise" is short-hand for "anything (be it an ink-deposit, a

pattern of light on a monitor, etc.) that might token an expression, the same qualification *mutatis mutandis* holding of related terms (e.g. "acoustic"). But, given some ink-deposit, *which* sentence-type it tokens is a function of what it means. A series of noises *acoustically* just like utterances of the English sentence "snow is white" could mean anything. Let L be a language such that, for some sentence S that belongs to L, S means chickens can fly" but utterances of S are *acoustically* indistinguishable from utterances of the English sentence "snow is white." Obviously utterances of S are not identical with utterances of "snow is white", and tokens of S are not identical with tokens of "snow is white," a corollary being that S and "snow is white" are different sentence-*types.* So sentence-types are not individuated by the physical properties of their tokens; and, what obviously follows, it isn't solely in virtue of a given object's physical (e.g. acoustical or morphological) properties that a given token is an instance of this as opposed to that sentence-type. Given some physical object x, is in virtue, not just of what x's physical properties are, but also of what it means, that x tokens this as opposed to that sentence-type. If a burst of noise acoustically just like a token of "snow is white" were produced a million years ago by some volcano, it would token no expression. Since meaning doesn't supervene on acoustics---since, in other words, two acoustically indistinguishable entities needn't coincide in meaning---no burst of noise *by itself* constitutes an expression-token.

[423] Which was enunciated by Frege (1892a). I don't know whether Frege was the first to state it.

[424] The order in which those parts occur is also relevant, that being why "the man Smith saw standing next to Jones" doesn't have the same referent as "the man Jones saw standing next to Smith."

[425] Frege (1892a) holds that, if ‹the phi› fails to refer, then ‹…the phi…› is without truth-value. For reasons that will presently become clear, this means that, even though Frege's analysis is in some crucial respects very different from Strawson's, the analyses coincide *in the respects that are relevant in this context*.

[426] Quine (1966) makes this point.

[427] "What if ‹x› is a fictitious or empty name?," it will be asked. Then ‹x is phi› is a pseudo-sentence or, at best, an open-sentence. "Wally Tweedbergler is an acrobat," I say. "Who does 'Wally Tweedbergler' refer to?," you ask. "No one," I respond. "I just made that name up." My utterance of "Wally Tweedbergler is an acrobat" is to be put in the same category as the non-sentence "fasdfdsaf is an acrobat" or the open- (and therefore non-)sentence "x is an acrobat." (Works of fiction are *ipso facto* prefaced by an implicit exhortation that one *suppose* the name-like expressions that one is about to encounter to be *actual* names. *The Muppets go to Hollywood* is implicitly prefaced by an exhortation that "Kermit" refer to a frog answering to a certain description. Although some authors (e.g. Meinong, Nathan Salmon) have taken the regressive and methodologically broken measure of saying that "Kermit is real," such chicanery is both unnecessary and useless. Useless because---news flash!---Kermit the frog isn't real, and "to say of what isn't, that it is, is to speak falsely," as Aristotle said. Unnecessary because, if we see *The Muppets go to Hollywood* as beginning with an (implicit) request that the viewer make certain suppositions, we can model the data without positing something we know not to exist.

See the paper titled "Mythical Objects" in Salmon (2005). See Chapter 8 of my (2010) *Analytic Philosophy* for a criticism of Salmon's argument. By Salmon's own admission, his argument, supposing it to establish that Kermit exists, also establishes that there is a fountain of youth, a rational square roots of 2, etc. (These are Salmon's own examples.) Salmon says that such objects *do* exist: they are, he says, non-spatiotemporal entities that humans being create (by producing works of fiction). First of all, non-spatiotemporal objects don't gush water and they don't give anybody eternal youth. Second, numbers are non-spatiotemporal to begin with. The number two is

the property of being a couple (or the class of all couples, or some such). So what is the rational square root of 2? Is it even *more* non-spatiotemporal than the number two? How is its being non-spatiotemporal supposed to make its existence possible?

[428] *Journal of Pragmatics*, July 2009.

[429] As McCready notes, Chomsky (1957) used this very sentence. As I recall, however, Chomsky uses it to show that sentences that hitherto haven't been used *do* have meanings. But if McCready is right to impute this view to Chomsky, then Chomsky is wrong, and for quite a variety of reasons, as we'll see. (Which is why I haven't bothered to see if, in fact, McCready's Chomsky-exegesis is correct.)

McCready's wording is ambiguous; and no sooner is it disambiguated than the problems with his contention begin to emerge. McCready says that CG has no "coherent meaning," which, he claims, decisively shows that *bona fide* sentences can be meaningless.

When McCready says that CG has no "coherent meaning," what is he saying? If he's saying that it *does* have a meaning, then his point defeats itself. So that's not what he's saying. He must be saying that, although its various parts have meanings, those various meanings fail to cohere into a single proposition (or, in any case, larger meaning of some kind or other). In other words, unless McCready's contention is very obviously false---unless, it's only a slight exaggeration to say, it's to the effect it is itself false----it must be to the effect that *nothing* is affirmed by (tokens of) CG. Which is what I will take McCready to be contending, both because, exegetically speaking, it clearly is what he's saying and, second, because a scholar of McCready's stature wouldn't make a brazenly self-defeating claim.

[430] My inclination is to think that, *at the level of semantics*, these sentences don't have logic or nomic force, and it is through implicatures, not explicature, that utterances of them transmit the logically/nomically pregnant claims just describes. I say this because those implicatures, if that's what they are, can be cancelled and are therefore, by Grice's (1975) plausible criterion, pragmatic epiphenomena, as opposed to semantic explicatures. Cf.

> "All metal objects expand when heated; but there's nothing inherent in what it is to be a metal object that disposes such objects to expand when heated. Long ago, hoping to mislead future scientists, some alien prankster, hoping to mislead future scientists destroyed all metal things that didn't expand when heated."

Most semanticists and pragmaticists accept Grice's (1975) views on cancellation. Wayne Davis (2003) rejects it. But he's an exception. Which doesn't mean he's wrong, of course---only that, by assuming *arguendo* the falsity of Davis' position, I'm arguing within a framework within which a lot of important linguistic and philosophical work is conducted and whose hypothetical correctness it is not unreasonable to assume.)

But, of course, it's a delicate question whether, in some contexts or some values of phi and psi, ‹all phi's and psi's› is, *at the level of literal meaning*, to the effect that a thing's a being phi is connected in some robust, not purely statistical way, to its being a psi.

[431] The distinction between the logically false and the logically ridiculous is a purely psychological distinction. Thanks to Gödel (1931), we know that

(i) "arithmetic can be formalized"

is false. But (i) isn't ridiculous, at least not to us. However,

(ii) "there are square circles"

***is* ridiculous.**

But to beings much more intelligent than ourselves, (i) wouild be quite as ridiculous as (ii), the same being true of:

(iii) "there are regular, 17-sided polygons"

and

(iv) "if the Theory of Types is true, any attempt to affirm it is false, wherefore it's false."

If, like McCready, we say that anything that, like (ii), "has no coherent meaning" is *ipso facto* devoid of meaning, then we're saying that each of (i) and (iii) and (iv) is devoid of meaning; and, by implication, we're saying that it is actually impossible to say anything that isn't logically, mathematically, philosophically, or otherwise non-empirically false. Which is itself non-empirically false. And ridiculous.

[432] Russell, B. (1940), p. 222. See also Scheffler, I. (1981), p. 159.

[433] Also, procrastination is a property of certain modes of conduct; and, being a property, it's non-spatiotemporal and therefore can no more be drunk than quadruplicity can drink. To be sure, *instances* o procrastination are spatiotemporal. (My putting off making my bed is something that happens, or doesn' happen as it were, in space-time.) But

(QP) "quadruplicity drinks ionstances of procrastination"

is false, since quadruplicity doesn't drink anything, as previously noted, and instances of procrastination not being liquids, cannot be drunk. So MS and QP are deeply idiotic, and therefore absurd in some sense of the word, but perfectly meaningful.

[434] Ryle is, I suspect, quite right to say that, because it is to be understood in organizational (functional), not (strictly) geographical terms, Oxford University could, without ceasing to be Oxford, relocate itself. There would have to be some sort of causal and functional nexus linking the pre- and pos repositioned entities for them to be constitutive of some one university. But that's consistent with Ryle's position and, in any case, isn't relevant here, since we'll grant Ryle's premise and show that, relative to Ryle's argument still fails.

[435] Carroll (2010).

[436] This sentence is borrowed from Hempel (2008).

[437] The brackets are meant to be quasi-quotes.

[438] Scheffler (1981), p. 159.

[439] "BW" is short for "boiling water."

[440] "GW" is short for "glubbified water."

[441] "BW" is short for "boiling water."

[442] "GW" is short for "glubbified water."

[443] Scheffler says that GW is meaningless *with respect to E*---where E is a hypothetical language such that no sentence S belongs to E if S is non-empirical (or non-tautologous). But Scheffler doesn't say, or acknowledge the possibility of saying, that GW is meaningless in some less conditional sense.

[444] There are obviously sentence-*types* that have no truth-conditions, e.g. "he laughed at her." In fact, in natural language, few if any sentence-types are entirely context-insensitive, given that few if any of them lack tense-markers. [444] But the logical positivists rightly disregarded the likes of "he ate it" and "she laughed derisively at her butler." And they were right to do so: when considered in a vacuum, "he laughed at her" can be treated as an open-sentence, the "he" and "her" being (restricted) free variables.

We'll see that there are *no* meaningless sentence-types or sentence-tokens. If E is an expression-type such that there is no rule saying what is meant by a given token t of E, then E is without meaning. If there is such a rule, then E is meaningful. If t is an a sentence-token that is assigned *truth-conditions*, even if t fails to have been assigned a proposition, t is still meaningful. If, while pointing at an empty space, I say "that elephant is a genius," my utterance is neither true nor false, since (if Strawson (1950) and Kaplan (1989) are right, as I think they are) the semantic rule for the corresponding sentence-type is:

> (SR) If, in context C, there is exactly one contextually salient elephant, and t is a token in C of "that elephant is a genius," t is true iff O is a genius and false if O is not a genius; if, in C, there is no contextually salient, t is without truth value.

But even though there is no elephant in the vicinity, t still has truth-*conditions*; that is why t---*that very token*---*could* have been true. It would have been true in a world otherwise exactly like our own in which there was a uniquely contextually salient elephant, which, moreover, was a genius. Also, in this world, when trying to determine whether a given token, such as t, is true, false or neither, we start our starting with its truth-conditions---with its *meaning*. (In what follows, "sentence-token" is to mean "sentence-token in the indicative mood.") For a sentence-token to be meaningful is not for there to be some proposition that it affirms (or even expresses); it is the semantic rules of the language to which it belongs to say what *must* be the case *if* that token is to be true. In other words, it is for that token to have truth-conditions. t does have truth-conditions. And, be it known, those conditions do not coincide with the semantic rule for SR. First of all, SR, unlike t, isn't true or false under *any* conditions, not even counterlogical ones. (Whenever I refer to hypothetical conditions, I am assuming that, in them, English is spoken. More generally, any mention, implicit or explicit, of a "possible world" or hypothetical scenario is followed by a (phonetically unrealized) token of "that is semantically just like our world, i.e. that is governed, in the same places, to the same degree, in the same way as our world." For it's obvious, and irrelevant in this context, that a token phonetically just like t could mean anything or nothing.

[445] See Ayer (1935, 1959), Bridgeman (1953), Carnap (1953a, 1953b, 1959a, 1959b), Hempel (1959), Neurath (1953), Reichenbach (1953), and Schlick (1959). For trenchant criticisms of the (logical positivist,

instrumentalist, hyper-empiricism) viewpoint espoused in these papers, see Gödel (1953) and Hempel (2008). The basis for much of Gödel (1953) is found in Gödel (1931). Also, See Kuczynski (2009), Chapter 1 for what is intended to be a thorough but compact analysis of logical positivism and its derivatives.

[446] I believe that the argument about to be put forth can be seen as a generalization of points made by Frege (1891).

[447] A technicality: the sentence "some people smoke and some people play squash" is not, technically, a quantified generalization, since the main connective is not "some people" (or any other quantifier), but rather "and." So it's a non-quantified molecular sentence in which two quantified sentences are embedded. In any case, such sentences are covered by what we'll say in Sections 3.2 and 3.3.

[448] According to Russell (1905, 1918) and Frege (1891), the logical form of "some person smokes" is "for some x, x is a person and x snores." The latter sentence can be interpreted as having the form ‹x has phi›. (Replace "x" with a noun denoting the property of being a person who smokes, and replace phi with an expression that denotes the property of being instantiated.) Thus, Frege's insight isn't necessarily represented as being to the effect "someone smokes" does *not* have the form ‹x has phi›; it can be interpreted as saying that it *does* have that form, *provided* that the grammatical subject and object ("someone" and "smokes," respectively) switch places (to put it loosely but suggestively) with logical subject and predicate ("the property of snoring" and "is instantiated," respectively.)

[449] See (2008), from which this example is borrowed.

[450] Sometimes the logical positivists described *actual* sentences as meaningless. For example, they claimed that statements about unobservables are meaningless *unless* construed as statements about observables (meter-readings, overt-behaviors). So "Sam is thinking about the real number system" is either meaningless or it's a statement about his overt behavior. (Cf. Ayer (1936, 1959), Carnap (1959b).) "Protons are more massive than electrons" is meaningless, unless it's an abbreviated way of speaking about experimental findings (cf. Carnap (1953) Bridgeman (1959)), the same being true of "the distance between the two ends of my swimming pool is smaller than the distance between Mercury and Jupiter" must (cf. Bridgeman (1959)).

Each of these views is obviously false. No statement about somebody's thought-process is equivalent with anything about his overt behavior. Any thought can lead to any body-movements, given the right circumstances. In general, anything can *cause* anything, given the right circumstances. Statements about effects aren't equivalent with statements about their causes. It would be absurd to say otherwise. The view in question says otherwise, it's absurd.

Any unobservable state of affairs can have any observable consequence, given the right circumstances. And, given the right circumstances, any attempt to determine the distance between two objects can have any observable consequence. So analogous of the just-stated argument establish the falsity of the other two just-mentioned views.

According to Ayer (1936), "Brutus Killed Caesar" is not about some past event, since that event cannot be observed, and, so far as it's about anything, is about historical records and other observable remnants of that event.

No sooner is Ayer's theory (AT) stated than refuted. In order to state AT, one must assume the reality of the very events that, if AT is right, cannot be meaningfully spoken about. If AT is correct, the term "past event" is a synonym of "the effects of past events." This means that AT is viciously circular in two ways. First, the meaning of "the effects of past events" is a function of the meaning "past events," so that the latter must have a meaning independently of the former. But let's set this aside, and let's concede to AT that these two expressions are synonymous. In that case, AT comes to this:

> AT*: "when meaningful, statements about the effects of past events in fact concern the effects of past events."

AT* is trivial and therefore doesn't have the same meaning as Ayer's demonstrably false claim that:

> AT# When meaningful, statements about past events actually concern the effects of past events.

Let E be some effect of Brutus' assassination of Caesar. The sentences ‹Brutus' assassination of Caesar caused E and therefore preceded› is true, whereas the statement ‹E caused E and therefore preceded it› is absurd. Therefore, those sentences don't say the same thing. Therefore, AT is false, given that, according to it, they *do* say the same thing.

[451] See Strawson (1950), Perry (1977), Barwise and Perry (1983), and Kaplan (1989). Strawson made most of the points that, in later years, would constitute the better part of the vigorous debate concerning the semantics of indexical (context-sensitive) sentences, even though, curiously, his name was almost never mentioned.

[452] Short for "that elephant."

[453] Short for "semantic (rule for that) elephant.")

[454] In this context, we're focusing on indicative sentences. But what we'll say is easily adapted to deal with non-indicative tokens---better adapted, as we'll see, than the contention than the contention we're now in the process of refuting, namely that the meanings of sentences are propositions, not truth-conditions, and therefore that, in an elephant free context, a token of TE is without meaning.

[455] Be it noted that, given any sentence-type that contains any context-sensitive component---an indexical (e.g. "now"), a "demonstrative (e.g. "that") or a tense-marker (e.g. "-ed")----that sentence-type's meaning cannot be a proposition. Given two tokens of "it's cold," both uttered in some one place but at different times, the one may express a true proposition while the other expresses a false one. Since no proposition is both true and false, those tokens express different propositions. The same thing *mutatis mutandis* holds with respect to *any* context-sensitive sentence-type.

Since every sentence of natural language contains a tense-marker, every such sentence is, at the strictly semantic level, context-sensitive. (In some cases, the context is so blatantly irrelevant that the tense-marker is innocuous (e.g. the tense-marker in "2 divided by itself equals 1") seems to do nothing. But supposing *per impossibile* that people believed (as some do) that mathematical reality changed, the tense-marker in that sentence would not be vestigial. And, were it to become widely accepted that mathematical reality was as changeable as the weather, even sentences like "2/1=1" and "2+2=4"---the formal analogues of "2 divided by itself equals 1" and "2 plus 2 equals 4"---might come to be seen not only as having tense-markers, but as having non-redundant ones. In any case, setting aside certain (numerically inclusive but) narrowly defined and singular classes of sentences, it's fair to say that most if not all natural-language sentences (sentence-types) contain context-sensitive components and that,

for this reason, the meanings of such types are rules of the previously-discussed kind.)

Kaplan (1989) refers to the meaning of a sentence-type as its "character"; and he refers to the proposition assigned to a given token of a given type as that token's "content." According to Kaplan, given some sentence type T, T's meaning is a rule that assigns *propositions* to tokens of T. I submit that T's meaning is a rule that assigns *truth-conditions* to tokens of T. I further submit that, for a sentence-token to be meaningful is for it to have truth-conditions and is *not* for there to be some proposition that it affirms. (A token's affirming some proposition is sufficient but not necessary for its being meaningful and, what is related, is a consequence, but not the essence, of its being meaningful.)

[456] Short for "semantic rule for woman/man."

[457] Short for "woman/man."

[458] Notice that (*) doesn't coincide with the meaning of K_1 (the *type* of which k_1 is a token). K_1's meaning is given by a rule that doesn't mention k_1 or any other specific token of K_1.

[459] In any case, this is the Strawson-Kaplan view, and for many reasons. (Cf. Strawson (1950), Kaplan (1989).) And it is a reasonable one.

[460] Advocates of CW (e.g. Russell (1908), Tarski (1930), and Kripke (1975)) take it to be to the discredit of the English language that the rules constitutive of it permit the construction of such a sentence. But, those rules being what they are, it is indeed a *bona fide* sentence-token of English.

[461] Not, *nota bene*, the expression that results when it is flanked in quotations; not, in general, any expression *referring to it*, whether homonymous with it or not; but that expression itself. In CW, tf *per se* occurs.

[462] Given *only* what t* means, no *empirical* work need be done to establish that it isn't true. But that doesn't mean that t* is false solely by virtue of what it means. The statement:

> (A) "Establishing that sentence-token s cannot be true involves no *empirical* knowledge (over and above that involved in understanding s to begin with)"

doesn't entail

> (B) "s's failure to be true is strictly a matter of language."

What does entail (B) isn't (A) alone, but rather (A) *conjoined* with:

> (C) "all truths are either spatiotemporal or semantic."

But (C) is false. Gödel (1931, 1953) made it clear that, even though it's a matter of semantics' that "3" refers to 3 and that "27" refers to 27, it's a matter of logic, not of convention, that, *given* those rules, "3" refers to the cube-root of the thing referred to by "27." In general, it's not a matter of convention what a given set of conventions entails. Nor, therefore, is it a matter of convention whether a given set of conventions is internally consistent, given that an inconsistent set of conventions is one that entails a contradiction.

[463] *A defense of this claim*: No series of noises is a sentence-*type*: an ink-deposit is, at most, a

sentence-token. (In this context, “burst of noise” is short-hand for “anything (be it an ink-deposit, a pattern of light on a monitor, etc.) that might token an expression, the same qualification *mutatis mutandis* holding of related terms (e.g. “acoustic”). But, given some ink-deposit, *which* sentence-type it tokens is a function of what it means. A series of noises *acoustically* just like utterances of the English sentence “snow is white” could mean anything. Let L be a language such that, for some sentence S that belongs to L, S means chickens can fly” but utterances of S are *acoustically* indistinguishable from utterances of the English sentence “snow is white.” Obviously utterances of S are not identical with utterances of “snow is white”, and tokens of S are not identical with tokens of “snow is white,” a corollary being that S and “snow is white” are different sentence-*types.* So sentence-types are not individuated by the physical properties of their tokens; and, what obviously follows, it isn’t solely in virtue of a given object’s physical (e.g. acoustical or morphological) properties that a given token is an instance of this as opposed to that sentence-type. Given some physical object x, is in virtue, not just of what x’s physical properties are, but also of what it means, that x tokens this as opposed to that sentence-type. If a burst of noise acoustically just like a token of “snow is white” were produced a million years ago by some volcano, it would token no expression. Since meaning doesn’t supervene on acoustics---since, in other words, two acoustically indistinguishable entities needn’t coincide in meaning---no burst of noise *by itself* constitutes an expression-token.

[464] Which was enunciated by Frege (1892). (Some of its more important consequences are clearly drawn in Barwise and Perry (1983).) I don’t know whether Frege was the first to state it. (I doubt that he was.)

[465] The order in which those parts occur is also relevant, that being why “the man Smith saw standing next to Jones” doesn’t have the same referent as “the man Jones saw standing next to Smith.”

[466] Frege (1892) holds that, if ‹the phi› fails to refer, then ‹…the phi…› is without truth-value. For reasons that will presently become clear, this means that, even though Frege’s analysis is in some crucial respects very different from Strawson’s, the analyses coincide *in the respects that are relevant in this context*. (See Dummett (1973) for an illuminating discussion of the ways in which Strawson’s analysis differs from Frege’s.)

[467] Quine (1966) makes this point.

[468] Some--e.g. Graham Priest (2006a, 2006b)--take it to show that there are true contradictions. Priest (correctly) believes that K_1 *is* meaningful and, on that basis, infer (understandably but erroneously) that it *ipso facto* affirms some proposition. Priest knows that *if* K_1 affirms a proposition, that proposition is both true and false. So Priest holds that some propositions are both true and false. Since, as we’ll now see, K_1 affirms nothing, this line of reasoning is spurious. (See Brown (2006) for a helpful discussion of Priest’s work.)

[469] “What if ‹x› is a fictitious or empty name?,” it will be asked. Then ‹x is phi› is a pseudo-sentence or, at best, an open-sentence. “Wally Tweedbergler is an acrobat,” I say. “Who does ‘Wally Tweedbergler’ refer to?,” you ask. “No one,” I respond. “I just made that name up.” My utterance of “Wally Tweedbergler is an acrobat” is to be put in the same category as the non-sentence “fasdfdsaf is an acrobat” or (best case scenario) the open- (and therefore non-)sentence “x is an acrobat.” (Works of fiction are *ipso facto* prefaced by an implicit exhortation that one *suppose* the name-like expressions that one is about to encounter to be *actual* names. *The Muppets go to Hollywood* is implicitly prefaced by an exhortation that “Kermit” refer to a frog answering to a certain description. Although some authors (e.g. Meinong, Nathan Salmon) have taken the regressive and methodologically broken measure of saying that “Kermit is real,” such chicanery is both

unnecessary and useless. Useless because---news flash!---Kermit the frog isn't real, and "to say of what isn't, that it is, is to speak falsely," as Aristotle said. Unnecessary because, if we see *The Muppets go to Hollywood* as beginning with an (implicit) request that the viewer make certain suppositions, we can model the data without positing something we know not to exist.

See the paper titled "Mythical Objects" in Salmon (2005). See Chapter 8 of my (2010) *Analytic Philosophy* for a criticism of Salmon's argument. By Salmon's own admission, his argument, supposing it to establish that Kermit exists, also establishes that there is a fountain of youth, a rational square roots of 2, etc. (These are Salmon's own examples.) Salmon says that such objects *do* exist: they are, he says, non-spatiotemporal entities that humans being create (by producing works of fiction). First of all, non-spatiotemporal objects don't gush water and they don't give anybody eternal youth. Second, numbers are non-spatiotemporal to begin with. The number two is the property of being a couple (or the class of all couples, or some such). So what is the rational square root of 2? Is it even *more* non-spatiotemporal than the number two? How is its being non-spatiotemporal supposed to make its existence possible?

[470] Let's start with a purely clarificatory point. The word "self" has two meanings or functions. Sometimes it can function as a garden-variety noun. That's how it's functioning in "Hume denied that he had a self." And sometimes "self" functions as a way of constructing 'terms of laziness,' to adapt an expression of Geach's, viz. expressions that are useful, because conducive to expedience, but technically eliminable. That's how it functions in

(J_1) John punched himself,

which is short for

(J_2) John punched John.

When functioning denotatively (as in "my self (in other words, my soul or psychological essence, or some such) is something that underlies and unifies the various fleeting occurrences and states constitutive of my consciousness") isn't eliminable (at least not in any relevant sense). But (so I am trying to show) when functioning *anaphorically*, it *is* eliminable (cf. J_1 and J_2). Henceforth, all references to the term "self" are to the anaphoric device.

Now the requisite qualifications of our contention (that "self" is always eliminable):

[471] Perry (1977) makes this point.

[472] All demonstratives are *indexicals*, but not *vice versa.* A demonstrative is an indexical that, in at least some contexts, needs to be accompanied by an act of demonstration (pointing or some such) in order for referent to be assigned to it. An indexical is an expression such (i) what a given one of its occurrences refers to is a function of the circumstances and such that (ii) it is *not* by virtue of that expression's being ambiguous that it has that property. "That guy" refers to different people on different occasions. But it isn't ambiguous. For there is some *one* semantic rule R such that, if a given token T of "that guy" refers to O, it is R that assigns O to T. The name "David," by contrast, *is* ambiguous; for refers to different people on different occasions *and* it does this by virtue of there being many different different semantic rules pairing off "David" with a person (or other object). So, supposing that $p_1 \ldots p_n$ are the people to whom "David" refers, existence, there is for any two numbers x and y ($1 \leq x \leq y \leq n$) a semantic rule to the effect that "David" (or tokens thereof) refer to p_x and a distinct semantic rule to the effect that "David" (same qualification) refers to p_y. An indexical $=_{DF}$An Expression E such that there is one semantic rule R according to which, given a situation S, and given an object O in S, there are some properties $p_1 \ldots p_n$ ($1 \leq n$) a given token of E

picks out O just in case O uniquely has all of $p_1 \ldots p_n$.

[473] Let IE (short for "indexical expression) be an arbitrary indexical. IE's "character" is the rule that assigns referents to IE's tokens. (So the character of "that duck" would be: given a context C, a token T of "that duck" refers to O exactly if, in C, O and O alone is a salient duck (with the qualification that a demonstrative act, e.g. one of pointing, may---or may not---be needed to render O salient). The "content" of a given (non-empty) token of that "that duck" is some duck. So if D is the duck picked out by the occurrence of "that duck" a token S of "that duck can fly," then S's literal meaning is (equivalent with) *O can fly*. So the indexical's character doesn't make it into the proposition of which the corresponding token is a constituent . What *does* make it in there is the object itself, *sans* any Fregean sense or anything of the sort. (I myself advocate the view that, if E directly refers to O, then "E can fly" encodes a proposition that has as a veritable constituent, not *O* per se, but rather *the property of being O*. This way, one can legitimately say, despite the fact that Socrates' no longer exists, a that "Socrates was wise" exactly if Socrates was wise, it being irrelevant who besides Socrates was wise and what properties besides wisdom Socrates had. So one is spared taking the desperate measure, urged by Nathan Salmon, of saying that "Socrates was wise" encodes a proposition, *albeit a non-existent one* and that "Socrates" has a referent, *albeit a non-existent one*. (If "Socrates" has a non-existent referent, there *ipso facto* exists nothing to which it refers and it has *no* referent. Salmon rejects the inference from "if E's referent is non-existent" to "E refers to nothing." In doing this, Salmon is mutilating the laws of logic so as to validate a contention internal to a small branch (semantics) of but one discipline (linguistics, philosophy of language). And, so I argue in __, ___, and ___, Salmon's position is unnecessary; there is no need to change the laws of logic, and certainly no need to change them in the way proposed by Salmon.

[474] Given only what I've said, this conclusion isn't actually warranted – it is, at most, suggested. But making this argument probative would take us astray. And it would also be unnecessary, since the point I'm trying to make (viz. that it isn't always obvious what is literally meant by an expression) is *itself* illustrated by the fact that it may be debated whether U1 and U2 have the same meaning.

[475] We are supposing that one knows st's meaning, not by virtue of simply being *told* it, but by virtue of one's commando of the English language.

[476] 'HIP' is short for 'Hesperus identical with Phosphorous.'

[477] A thorough explanation of this is given in Chapter__

[478] In any case, here's a summary: In its crudest form logical positivism (LP) is the doctrine that:

(1) A statement S is meaningful iff it's either definitionally true or such that there are observations (possible if not actual) that would completely verify it.

Three problems with (1`): Problem 1: It's up to us what are definitions are, but not whether they're consistent with one another or what they're consequences are. (It's up to us what is meant by "rational number," "square roote," "2," etc.; but it's a non-conventional consequence of those conventions that "2 a rational square root is false" Problem: Since it expresses a normative, as opposed to strictly claim, (1) is itself neither empirically verifiable; and (1) is clearly not a tautology. (3) "All metal expands when heated" is meaningful but not verifiable (unless the concept of verification is so broadened that it may involve the use of principles that, being neither tautologous nor capable of strictly sensory certification, are LP-unfriendly). The principle that

(2) S is meaningful iff either a definitional truth or if empirically falsifiable

suffers from the same problems (*mutatis mutandis*) as (1). Which is to be expected since its really a form of verificationism: "There is no gold ball weighing exactly 2.38766 kgs" (not falsifiable)≡ "Any given object either isn't a gold ball or it doesn't exactly weigh 2.38766 kgs" (not verifiable). Plus, (2) is no better than (1) at dealing with the fact that it's not a matter of definition (or convention) whether a plurality of definitions (conventions) are consistent or what they're consequences are.

This last point same also holds the final, most liberal form of LP, namely:

(3) S is meaningful if a tautology or if capable of being supported by sensory observation.

The problem with (3) is that, given any collection of data $D_1 \ldots D_n$, and given any prediction P, no matter how obviously false, there is some set of predicates $p_1 \ldots p_m$ that, if that data is described in terms of those predicates, acceptance of P is *de riguer*. For example, let an "iguanacar" be anything that is a car before the year 2020 AD and is an iguana thereafter. All cars thus far have been iguanacars. So there is as much statistical support for the predication that cars will become iguanas in a few years as there is that they won't. The one prediction is reasonably whereas the other is not. Whatever the principle is that we countenance that leads us to reject the one inference, along with the bent predictates associated with it, that principle cannot be justified on strictly sensory grounds, since an inductive justification of that principle can be perverted by contriving bent predicates that warrant acceptance of a clearly spurious principle. So our predictive practices are conditioned by an acceptance of principles for which there is no strictly perceptual justification. (See See Hempel (1950), Hempel (1966); see also Goodman (1965) for an argument similar to the "iguanacar" argument just presented.)

[479] The principles underlying the sentence-reparsing that I do in this paper are ultimately due to Frege (1892b) but are made (relatively explicit) for the first time in Montague (1974).

[480] Martin Heidegger did in fact say this. But the fact that one person said it doesn't mean that the logical positivists were not attacking a straw man.

[481] Of course, that's not really a sentence; it's a sentence-schema. "For all statements P and Q, if P, then either P or Q" is a sentence proper, but it's not formally true, since---unless "all" is circularly considered to be such that, in virtue of containing it where an otherwise identical sentence contains "for no x"---it has the same form as "for no statements P and Q, if P then P or Q," which is false.

[482] See Barwise and Etchemendy (1987).

[483] See Russell (1908).

[484] See Priest (1995).

[485] For reasons to be discussed in Chapter 7 of my book *Analytic Philosophy*, and vaguely adumbrated in Chapter 2 of this book, this means that statements about numbers must either be of Order 3 or of some higher order. Basically, numbers are properties of properties. (The number two is the property had in common by all and only those sets *S* such that, for some *x* and *y*, $x \neq y$ and *x* and *y* belong to *S* and anything *z* that belongs to *S* is identical with either *x* or *y*. Since the number two is a second-order property, statements *about* the number two are of Type_3.)

[486] Quine (1966) makes this point in the paper entitled “the Ways of Paradox,” which is also the title of the book in which that paper occurs.

[487] Short for “alternate interpretation of RP.”

[488] What follows is a repetition of the main points concerning context-sensitivity made in Chapter 3. I am doing this to make this chapter as self-contained as possible. In any case, the rest of Section 3.2 may be skipped by those who have absorbed that material.

[489] Parsons (2000) is an example of such a philosopher. Keefe (1997) contains many articles, by various philosophers, advocating the view that there is objective vagueness. In his unpublished (UCLA) dissertation, which was supervised by Kit Fine (himself an advocate of the view that there is objective vagueness), Paul Hovda (2000) argues that there is objective vagueness. (His argument is simply that it’s implausible to suppose that the world is articulated into objects all of which have sharp edges. He doesn’t consider the argument presented here.)

[490] “OV” stands for “objective vagueness.”

[491] “SB” is short for “Smith is bald.”

[492] “ER” is short for “entailment relations.”

[493] Or, more precisely, any given one of the infinitely many possible ways.

[494] Examples of predicates are “tall,” “slow,” and “prime.” Predicates are typically defined as “expressions that pick out properties.” But this definition, though decent enough for some purposes, won’t do as it stands. For there are non-predicates; e.g., “the property of being tall,” that pick out properties. It is not easy to come up with a viable definition of predicate; for any such definition will presuppose the truth of substantive theoretical claims. In any case, here is (what I believe to be) a viable definition of that term. A predicate is an expression *E* such that *E* is elliptical for an open sentence that has the form “*x* is an instance of the property of being *P*” and that, because it has the form, yields an actual sentence when the variable is replaced with either a referring term (e.g., “John”) or a quantifier (e.g., “some penguin”). So “tall,” according to this view, is a predicate because it is elliptical for (a grammatical variant of) the expression “*x* has the property of tallness,” which, when the variable is replaced with a referring term or quantifier, yields a *bona fide* sentence (e.g., “John has the property of tallness"). The resulting sentence is abbreviated into one that, at the level of orthography, has the form “*x* is *P*.” (So, for example, “John has the property of tallness” is abbreviated into “John is tall.”) This theory—though artificial in some ways---does explain how exactly predicates (e.g., “tall”) differ from the corresponding noun-phrases (“the property of tallness”) and, so far as I know, is otherwise consistent with the data.

[495] Quine (1953) uses a similar example to illustrate a different point.

[496] I say “relatively” because I don’t wish to rule out that if, given some multiplicity of grains of sand, the distance between any given two of those grains of sand is sufficiently small, that multiplicity could still constitute a heap. If the distance between any two grains is small enough that, as far as the physical behavior of that heap is concerned, those two grains *might as well be contiguous*, then that multiplicity would constitute a heap. That said, to my knowledge, it isn’t physically possible for a multiplicity of non-contiguous grains of sand to have the dynamic properties of a multiplicity of contiguous grains of sand. More importantly, it could be argued that, if a multiplicity of grains of sand *behaved* as though its constituents were contiguous with one another, then they *ipso facto* would be contiguous with another---that the forces binding the grains of sand together, whereby those grains of

sand jointly constitute a single object, *make* those grains of sand be contiguous, at least by one legitimate metric.

[497] H: The hypothesis that all phi's are psi's. *First argument*. Premise: Exactly 1,000,000 objects are known to be a phi; it turns out that each is a psi *and* also that each is otherwise exactly like each of the remaining 999,999. Conclusion: Nothing can be concluded as to how likely it is that H is true.

Second argument. Premise: Exactly ten objects are known to be phi's; it turns out that each is a psi and also that no two of those objects are otherwise at all similar. Conclusion: H is almost certainly correct.

[498] "SB" is short for "stilts beanie"

[499] Goodman (1983).

[500] "GA" is short for "Goodman's argument."

[501] Short for objection to the author's use of Goodman's argument.

[JK1]

[JK2]Close artgument?

[JK3]Or, at least is typically used to say something non-trivial.